GLOBAL-IN-TIME WEAK MEASURE SOLUTIONS AND FINITE-TIME AGGREGATION FOR NONLOCAL INTERACTION EQUATIONS

J. A. CARRILLO[1], M. DIFRANCESCO[2], A. FIGALLI[3], T. LAURENT[4] AND D. SLEPČEV[5]

Abstract. In this paper we provide a well-posedness theory for weak measure solutions of the Cauchy problem for a family of nonlocal interaction equations. These equations are continuum models for interacting particle systems with attractive/repulsive pairwise interaction potentials. The main phenomenon of interest is that, even with smooth initial data, the solutions can concentrate mass in finite time. We develop an existence theory that enables one to go beyond the blow-up time in classical norms and allows for solutions to form atomic parts of the measure in finite time. The weak measure solutions are shown to be unique and exist globally in time. Moreover, in the case of sufficiently attractive potentials, we show the finite time total collapse of the solution onto a single point for compactly supported initial measures. Our approach is based on the theory of gradient flows in the space of probability measures endowed with the Wasserstein metric. In addition to classical tools, we exploit the stability of the flow with respect to the transportation distance to greatly simplify many problems by reducing them to questions about particle approximations.

Keywords: well-posedness for measure solutions, gradient flows, optimal transport, nonlocal interactions, finite time blow-up, particle approximation.

AMS Classification: 35B40, 45K05, 49K20, 92DXX.

1. Introduction

We consider a mass distribution of particles, $\mu \geq 0$, interacting under a continuous *interaction potential*, W. The associated *interaction energy* is defined as

$$\mathcal{W}[\mu] := \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\,d\mu(x)\,d\mu(y). \tag{1.1}$$

Our paper is devoted to the class of continuity equations of the form

$$\frac{\partial\mu}{\partial t} = \operatorname{div}\left[\left(\nabla\frac{\delta\mathcal{W}}{\delta\mu}\right)\mu\right] = \operatorname{div}\left[(\nabla W * \mu)\mu\right] \qquad x\in\mathbb{R}^d\,, t>0. \tag{1.2}$$

The equation is typically coupled with an initial datum

$$\mu(0) = \mu_0. \tag{1.3}$$

The velocity field in the continuity equation, $-(\nabla W * \mu)(t,x)$, represents the combined contributions, at the point x, of the interaction through the potential W with particles at all other points.

Date: November 22, 2010.

The choice of W depends on the phenomenon studied. For instance in population dynamics, one is interested in the description of the evolution of a density of individuals. Very often the interaction between two individuals only depends on the distance between them. This suggests a choice of W as a *radial* function, i.e. $W(x) = w(|x|)$. Moreover, a choice of w such that $w'(r) > 0$ corresponds to an *attractive* force among the particles (or individuals), whereas $w'(r) < 0$ models a *repulsive* force.

Equation (1.2) arises in several applications in physics and biology. Simplified inelastic interaction models for granular media were considered in [4, 18] with $W = |x|^3/3$ and [43, 28] with $W = |x|^\alpha$, $\alpha > 1$. Such models usually lead to convex attractive potentials.

Mathematical modeling of the collective behavior of individuals, such as swarming, has also been treated by continuum models steaming from discrete particle models [33, 14, 41, 34, 42, 15, 36, 13, 21, 23, 16, 17]. Typical examples of interaction potentials appearing in these works are the attractive Morse potential $W(x) = -e^{-|x|}$, attractive-repulsive Morse potentials $W(x) = -C_a e^{-|x|/\ell_a} + C_r e^{-|x|/\ell_r}$, $W(x) = -e^{-|x|^2}$, $W(x) = -C_a e^{-|x|^2/\ell_a} + C_r e^{-|x|^2/\ell_r}$, or W being the characteristic function of a set in $\mathbb{R}^d$. A major issue is the possibility of a *finite time blow-up* of initially regular solutions, which occurs when w is attractive enough near $r = 0$. In particular, the solution can aggregate (collapse) part (or all) of its mass to a point in finite time. Blow-up producing potentials feature a suitable singularity in their second derivative at $r = 0$. Typically, the potential is of the form $W(x) \approx |x|^{1+\alpha}$ with $0 \le \alpha < 1$, see [27, 7, 6, 5] in case of the Lipschitz singularity. Related questions with diffusion added to the system have been tackled in [9, 29, 30, 31].

Finally, another source of models with interaction potential appear in the modeling of cell movement by chemotaxis. In fact, the classical Patlak-Keller-Segel [38, 26] system, see [12, 10, 11], corresponds to the choice of the Newtonian potential in $\mathbb{R}^2$ as interaction, $W = \frac{1}{2\pi} \log |x|$, with linear diffusion. In the case without diffusion, a notion of weak measure solutions was introduced in [39] for which the author proved global-in-time existence, although uniqueness is lacking.

Given a continuous potential W, thanks to the structure of (1.2), we can assume without loss of generality that the following basic assumption holds:

(NL0) W is continuous, $W(x) = W(-x)$, and $W(0) = 0$.

Moreover, the potentials considered in this paper will also satisfy the following assumptions:

(NL1) W is λ–convex for some $\lambda \le 0$, i.e. $W(x) - \frac{\lambda}{2}|x|^2$ is convex.
(NL2) There exists a constant $C > 0$ such that

$$W(z) \le C(1 + |z|^2), \qquad \text{for all } \ z \in \mathbb{R}^d.$$

(NL3) $W \in C^1(\mathbb{R}^d \setminus \{0\})$.

We will say that the potential is a *pointy potential* if it satisfies **(NL0)**-**(NL3)** and it has a Lipschitz singularity at the origin. If in addition, the potential is continuously differentiable at the origin, we will speak about a *C^1-potential.* If 0 is a local minimum of the potential W, we will say that the potential is *locally attractive.* Note that any potential which is λ-convex for a positive λ is also λ-convex for $\lambda = 0$ and thus satisfies assumption **(NL1)**.

Remark 1.1. Assumptions **(NL0)-(NL1)** imply that

$$W(x) \geq \frac{\lambda}{2}|x|^2, \tag{1.4}$$

since $0 \in \partial W(0)$ and $W(0) = 0$. Hypotheses **(NL1)-(NL3)** imply a growth control on the gradient of W. More precisely, using the convexity of $x \mapsto \tilde{W}(x) := W(x) - \frac{\lambda}{2}|x|^2$ and the quadratic growth of $W(x)$, there exists $K > 0$ such that

$$\nabla \tilde{W}(x) \cdot p \leq \tilde{W}(x+p) - \tilde{W}(x) \leq K(1 + |x|^2 + |p|^2)$$

for any $x \neq 0$. Now, taking the supremum among all vectors p such that $|p| = \max\{|x|, 1\}$, we get $|\nabla \tilde{W}(x)| \leq K(2 + 2|x|)$ from which

$$|\nabla W(x)| \leq 2K + (2K + |\lambda|)\, |x|. \tag{1.5}$$

Let us also remark that **(NL1)** together with **(NL3)** imply that if the potential is not differentiable at the origin, then it has at most a Lipschitz singularity at the origin. Examples of locally attractive potentials neither pointy nor smooth are the ones with a local behavior at the origin like $|x|^{1+\alpha}$, with $0 < \alpha < 1$.

The first problem we treat in this paper is to give a well-posedness theory of weak measure solutions in the case of pointy potentials. Due to the possible concentration of solutions in a finite time, one has to allow for a concept of weak solution in a (nonnegative) measure sense. Our work fills in an important gap in the present studies of the equation. Simplistically speaking: on the one hand, [2, 3] provide a good theory for weak measure solutions for potentials which are either smooth or do not produce blow-up in finite time. Indeed, when solutions concentrate and the potential is not everywhere differentiable, this is the first paper where one is able to characterize the subdifferential of $\mathcal{W}$, see Proposition 2.2. On the other hand, in the works that study potentials that do produce blow-up [27, 7, 6, 5] the notion of the solution breaks down at the blow-up time.

Before discussing the main results of this work, we introduce the concept of weak measure solution to (1.2). A natural way to introduce a concept of weak measure solution is to work in the space $\mathcal{P}(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$. Since the class of equations described here does not feature mass–threshold phenomena, we can normalize the mass to 1 without loss of generality, due to the following invariance of the equation: if $\mu(t)$ is a solution, so is $M\mu(Mt)$ for all $M > 0$. Following the approach developed in [2, 3], we shall consider weak measure solutions which additionally belong to the metric space

$$\mathcal{P}_2(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^2 \, d\mu(x) < +\infty \right\}$$

of probability measures with finite second moment, endowed with the 2–*Wasserstein distance* d_W; see the next section.

Definition 1.2. *A locally absolutely continuous curve* $\mu : [0, +\infty) \ni t \mapsto \mathcal{P}_2(\mathbb{R}^d)$ *is said to be a* weak measure solution *to* (1.2) *with initial datum* $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ *if* $\partial^0 W * \mu$ *belongs to* $L^1_{loc}([0, +\infty); L^2(\mu(t)))$ *and*

$$\int_0^{+\infty} \int_{\mathbb{R}^d} \frac{\partial \varphi}{\partial t}(x,t)\, d\mu(t)(x)\, dt + \int_{\mathbb{R}^d} \varphi(x,0)\, d\mu_0(x) = \int_0^{+\infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} \nabla \varphi(x,t) \cdot \partial^0 W(x-y)\, d\mu(t)(x)\, d\mu(t)(y)\, dt, \tag{1.6}$$

for all test functions $\varphi \in C_c^\infty([0,+\infty) \times \mathbb{R}^d)$.

In this definition, $\partial^0 W(x)$ denotes the element of minimal norm in the subdifferential of W at x. In particular, thanks to our assumptions on W the formula

$$(\partial^0 W * \mu)(x) = \int_{y \neq x} \nabla W(x-y)\, d\mu(y)$$

holds. Here, the absolute continuity of the curve of measures means that its metric derivative is integrable, see next section. Let us point out that, as a consequence of (1.5), $\partial^0 W * \mu \in L^2(\mu)$ for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

The main idea to construct weak measure solutions to (1.2) is to use the interpretation of these equations as gradient flows in the space $\mathcal{P}_2(\mathbb{R}^d)$ of the interaction potential functional (1.1) with respect to the transport distance d_W. Such an interpretation turns out to be extremely well-adapted to proving uniqueness and stability results for *gradient flow solutions* compared to other strategies. This basic intuitive idea, introduced in [37] for the porous medium equation and generalized to a wide class of equations in [19], was made completely rigorous for a large class of equations in [2, 3] including some particular instances of (1.2). For gradient flow solutions we are able to obtain the existence, uniqueness and d_W-stability. Let us point out that gradient flow solutions are eventually shown to be equivalent to weak measure solutions.

Let us remark that the well-posedness theory of gradient flow solutions in the space of probability measures is developed in [2, 3] for λ-convex potentials. However, a characterization of the subdifferential of the interaction functional $\mathcal{W}$ is provided only when the potential W is C^1. Here, we mainly focus on generalizing this theory to allow Lipschitz singularities at the origin. In the case of not C^1-potentials satisfying **(NL0)**-**(NL3)**, the technical point to deal with is the characterization of the subdifferential and its element of minimal norm. Moreover, we generalize this gradient flow theory allowing a negative quadratic behaviour at infinity. This fact introduces certain technical difficulties at the level of coercivity and lower semicontinuity of the functional defining the variational scheme. The well-posedness theory of gradient flow solutions is the goal of Section 2.

One of the key properties of the constructed solutions is the stability with respect to d_W: given two gradient flow solutions $\mu^1(t)$ and $\mu^2(t)$,

$$d_W(\mu^1(t), \mu^2(t)) \le e^{-\lambda t}\, d_W(\mu_0^1, \mu_0^2)$$

for all $t \ge 0$. If $\lambda > 0$ the above estimate still holds provided that the initial measures have the same center of mass, see Remark 2.14. The above stability estimate is not only useful for showing uniqueness but it is mainly a tool for approximating general solutions by particle ones. In fact, the previous estimate can be considered as a proof of the convergence of the continuous particle method for this equation on bounded time intervals. This is very much in the spirit of early works in the convergence of particle approximations to Vlasov-type equations in kinetic theory [22, 35, 40].

Let us finally mention that it is not difficult to check that weak-L^p solutions with initial data in $L^1_+ \cap L^p(\mathbb{R}^d)$ with finite second moment constructed in [5, 20, 8] are also weak measure solutions in the sense of Definition 1.2 up to their maximal existence time, see Remark 2.15.

Section 3 is devoted to show qualitative properties of the approximate solutions obtained by the variational scheme as in [25]. More precisely, we prove that particles remain particles at the level of a discrete variational scheme, provided the time step is small enough. In particular, this shows that the gradient flow solution starting from a finite number of particles remains at any time a finite number of particles, whose positions are determined by an ODE system. Although one can check directly via the solution concepts that such construction provides the solution for a finite number of particles, it is quite interesting to prove this property directly at the variational scheme level, as it is shown by its suitability as a numerical scheme.

Section 4 is devoted to the question of finite-time blow-up of solutions. For a radially symmetric attractive potential, i.e. $W(x) = w(|x|)$, $w'(r) > 0$ for $r > 0$, the number

$$T(\varepsilon_1) := \int_0^{\varepsilon_1} \frac{dr}{w'(r)}, \qquad \varepsilon_1 > 0 \tag{1.7}$$

can be thought as the time it takes for a particle obeying the ODE $\dot{X} = -\nabla W(X)$ to reach the origin if it starts at a distance ε_1 from the origin. This number quantifies the attractive strength of the potential: the smaller $T(\varepsilon_1)$ is, the more attractive the potential is. It was shown in [6, 7, 8] that if $T(\varepsilon_1) = +\infty$ for some (or equivalently for all) $\varepsilon_1 > 0$, then solutions of (1.2) starting with initial data in L^p will stay in L^p for all time, whereas if $T(\varepsilon_1) < +\infty$ for some $\varepsilon_1 > 0$, then compactly supported solutions will leave L^p in finite time (this result holds in the class of potentials which does not oscillate pathologically around the origin). Here, thanks to our developed existence theory, we are able to obtain further understanding of the large time behavior of the solutions: loosely speaking, we prove that if the potential is attractive enough (i.e. $T(\varepsilon_1) < +\infty$ for some $\varepsilon_1 > 0$) then solutions of (1.2) starting with measure initial data will concentrate to a single Delta Dirac in finite time. We refer to this phenomena as finite time total collapse.

We will say that W is an *attractive non-Osgood potential* if in addition to **(NL0)-(NL3)**, it satisfies the finite time blow-up condition:

> **(NL-FTBU)** W is radial, i.e. $W(x) = w(|x|)$, $W \in C^2(\mathbb{R}^d \setminus \{0\})$ with $w'(r) > 0$ for $r > 0$ and satisfying the following monotonicity condition: either **(a)** $w'(0^+) > 0$, or **(b)** $w'(0^+) = 0$ with $w''(r)$ monotone decreasing on an interval $(0, \varepsilon_0)$. Moreover, the potential satisfies the integrability condition
>
> $$\int_0^{\varepsilon_1} \frac{1}{w'(r)} dr < +\infty, \qquad \text{for some } \varepsilon_1 > 0. \tag{1.8}$$

Let us point out that the condition of monotonicity of $w''(r)$ is not too restrictive. It is actually automatically satisfied by any potential which satisfies (1.8) and whose second derivative does not oscillate badly at the origin, as in [6, 7] (more comments on this assumption are done in Section 4). Examples of this type of potentials are the ones having a local behavior at the origin like $w'(r) \simeq r^\alpha$ with $0 \le \alpha < 1$ or $w'(r) \simeq r \log^2 r$.

The proof of finite-time total collapse of solutions for attractive non-Osgood potentials is based on showing a finite-time total collapse result for the particles approximation independent of the number of particles, but possibly depending on the initial support. This fact, together with the convergence of the particle approximation, leads to the finite-time aggregation onto a single particle with the total mass of the system. This is the main technical novelty of our approach to blow-up.

It is worthwhile to remark on how our finite time total collapse result relates to previous works on finite time blow-up of weak-L^p solutions [7, 5, 8]. It was shown in [8] that a weak-L^p solution will exist as long as its L^p-norm is bounded. Since weak-L^p solutions agree with the weak measure solutions for as long as weak-L^p solutions exist, if the finite time collapse occurs then there exists a time T^* such that the L^p-norm of the density of the measure $\mu(t)$ goes to infinity as $t \to T^*$. That is, the finite time collapse of weak measure solutions implies finite time blow-up in the L^p-norm for weak-L^p solutions. In this way, we recover the results of finite time blow-up for more restrictive potentials W obtained in [6] under the condition **(NL-FTBU)**. We emphasize that the condition **(NL-FTBU)** implies both the finite time blow-up in L^p and the finite time collapse. Let us also point out that, even if we extend the notion of solution in a unique way after any L^p blow-up time, we are not able with this strategy to characterize the typical profile of L^∞ or L^p blow-up. We refer the reader to [24] for a numerical study of this question. Let us remark that the blow-up of the solution in L^p-norms will in general happen before the total aggregation/collapse onto a single point. The transition from the first L^∞ blow-up to the total collapse can be very complicated. For instance one could have multiple points of aggregation onto Dirac deltas interacting between them and with smooth parts of the measure in a challenging evolution before the total aggregation onto a single point. As explained in Section 4, as a consequence of the strategy of proof for the finite time total collapse, we can exhibit the appearance of multiple collapses into different Dirac deltas which eventually will collapse all together, see Proposition 4.6. This also shows that generically any L^p blow-up will happen before the total collapse time except for very particular initial symmetric distributions. This complex behavior was already encountered in [39] in the case of the chemotaxis model without diffusion, but his notion of solution lacks of uniqueness and stability. Many problems on the details of the blow-up in (1.2) and the interaction of delta masses with surrounding absolutely-continuous-measure part remain open.

2. The Jordan–Kinderlehrer–Otto (JKO) scheme

In this section we develop the existence theory for measure–valued solutions in the sense of Definition 1.2 by following the set up developed in [2]. A natural choice of a space of measures where to develop such a theory is the space $\mathcal{P}_2(\mathbb{R}^d)$ endowed with the *Wasserstein* distance

$$d_W(\mu,\nu) := \left[\min\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\gamma(x,y) :\ \gamma\in\Gamma(\mu,\nu)\right\}\right]^{1/2}, \tag{2.1}$$

where the set $\Gamma(\mu,\nu)$ of *transport plans* between μ and ν is defined by

$$\Gamma(\mu,\nu) := \left\{\gamma\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d) : (\pi_1)_\#\gamma=\mu \text{ and } (\pi_2)_\#\gamma=\nu\right\}$$

with $\pi_1(x,y)=x$ and $\pi_2(x,y)=y$, that is,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(x)d\gamma=\int_{\mathbb{R}^d}\phi(x)d\mu,\quad \int_{\mathbb{R}^d\times\mathbb{R}^d}\phi(y)d\gamma=\int_{\mathbb{R}^d}\phi(y)d\nu,\quad \text{for all}\,\phi\in C_b(\mathbb{R}^d).$$

The space $(\mathcal{P}_2(\mathbb{R}^d), d_W)$ is a complete metric space [44, 2]. The standard theory of optimal transportation [2, 44] provides the existence of an optimal transport plan for the variational problem (2.1), i.e. there exists $\gamma_o\in\Gamma(\mu,\nu)$ such that

$$d_W^2(\mu,\nu)=\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2 d\gamma_o(x,y). \tag{2.2}$$

The set of all the *optimal plans* γ_o satisfying (2.2) is denoted by $\Gamma_o(\mu,\nu)$.

We recall that the *interaction energy* $\mathcal{W} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is defined as follows:

$$\mathcal{W}[\mu] := \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y) d\mu(x) d\mu(y). \tag{2.3}$$

Note that $\mathcal{W}$ is well-defined on $\mathcal{P}_2(\mathbb{R}^d)$ due to assumptions **(NL1)-(NL2)**, which provide suitable control of the integral at infinity. Also, the continuity of W ensures the well-posedness of $\mathcal{W}$ on singular measures.

Following [2], we shall first address the problem of the existence of a *curve of maximal slope* for the functional $\mathcal{W}$. For this purpose, let us introduce some definitions. The *slope of* $\mathcal{W}$ is defined as:

$$|\partial \mathcal{W}|[\mu] := \limsup_{\nu \to \mu} \frac{(\mathcal{W}[\mu] - \mathcal{W}[\nu])^+}{d_W(\mu, \nu)}, \tag{2.4}$$

where $u^+ := \max\{u, 0\}$. Given an absolutely continuous curve $[0, T] \ni t \mapsto \mu(t) \in \mathcal{P}_2(\mathbb{R}^d)$, its *metric derivative* is:

$$|\mu'|(t) := \limsup_{s \to t} \frac{d_W(\mu(s), \mu(t))}{|s-t|}. \tag{2.5}$$

Finally, we recall the definition of a curve of maximal slope for the functional $\mathcal{W}$. With the notation in [2], such a notion is referred to as a "curve of maximal slope with respect to $|\partial \mathcal{W}|$".

Definition 2.1. *A locally absolutely continuous curve* $[0, T] \ni t \mapsto \mu(t) \in \mathcal{P}_2(\mathbb{R}^d)$ *is a* curve of maximal slope *for the functional* $\mathcal{W}$ *if* $t \mapsto \mathcal{W}[\mu(t)]$ *is an absolutely continuous function, and the following inequality holds for every* $0 \le s \le t \le T$*:*

$$\frac{1}{2} \int_s^t |\mu'|^2(r)\, dr + \frac{1}{2} \int_s^t |\partial \mathcal{W}|^2[\mu(r)]\, dr \le \mathcal{W}[\mu(s)] - \mathcal{W}[\mu(t)]. \tag{2.6}$$

The notion of solutions provided in Definition 2.1 is purely metric (see [2, Part I]). We shall improve this notion of solution (in the spirit of [2, Part II]) to a solution in the "gradient flow" sense in Subsection 2.3.

The inequality (2.6), which defines the notion of a curve of maximal slope, is better understood after providing a representation formula for the slope $|\partial \mathcal{W}|$ in terms of an integral norm of a vector field involving the "gradient" of W, or rather its minimal subdifferential $\partial^0 W$. Moreover, the metric derivative $|\mu'|$ should be interpreted in a "length space" sense, which accounts for the metric space $\mathcal{P}_2(\mathbb{R}^d)$ being endowed with a kind of Riemannian structure, first introduced in [37] and then proven rigorously in [2]. For the sake of clarity, let us briefly recall this framework (see [2, Chapter 8] for further details).

Given a measure $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, the tangent space $Tan_\mu \mathcal{P}_2(\mathbb{R}^d)$ to $\mathcal{P}_2(\mathbb{R}^d)$ at μ is the closed vector subspace of $L^2(\mu)$ given by

$$Tan_\mu \mathcal{P}_2(\mathbb{R}^d) := \overline{\{\nabla \phi \ : \ \phi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\mu)}.$$

Moreover, given an absolutely continuous curve $t \mapsto \mu(t) \in \mathcal{P}_2(\mathbb{R}^d)$, the "tangent vectors" to $\mu(t)$ can be identified as elements of the set of vector fields $v(t)$ solving the continuity equation

$$\partial_t \mu(t) + \nabla \cdot (v(t) \mu(t)) = 0 \tag{2.7}$$

in the sense of distributions. Among all the possible velocity fields $v(t)$ solving (2.7), as a consequence of [2, Theorem 8.3.1], there is one with minimal $L^2(\mu(t))$-norm, equal to the metric derivative of $\mu(t)$. Therefore, we have the following representation formula for $|\mu'|(t)$: for a.e. $t \in (0,T)$,

$$|\mu'|(t) = \min\left\{ \|v(t)\|_{L^2(\mu(t))} : v(t) \text{ solves (2.7) in the sense of distributions}\right\}.$$

More precisely, for every solution $v(t)$ of (2.7) the inequality $|\mu'|(t) \le \|v(t)\|_{L^2(\mu(t))}$ holds at a.e. $t \in (0,T)$, and there exists a "unique" solution of (2.7) for which equality holds a.e. on $(0,T)$, see [2, Theorem 8.3.1 and Proposition 8.4.5]. We recall here the upper bound

$$\limsup_{\varepsilon \searrow 0} \frac{d_W((id+\varepsilon\xi)_\# \mu, \mu)}{\varepsilon} \le \|\xi\|_{L^2(d\mu)} \tag{2.8}$$

which follows immediately from the trivial inequality

$$W_2(S_\# \mu, T_\# \mu) \le \|S - T\|_{L^2(d\mu)}.$$

As for the slope $|\partial\mathcal{W}|$ of the functional $\mathcal{W}$ (similarly to the classical subdifferential calculus in Hilbert spaces), it can be written as

$$|\partial\mathcal{W}|(\mu) = \min\left\{ \|w\|_{L^2(\mu)} : \ w \in \partial\mathcal{W}(\mu)\right\},$$

where $\partial\mathcal{W}(\mu)$ is the (possibly multivalued) *subdifferential* of $\mathcal{W}$ at the measure μ. The definition of subdifferential of a functional $\mathcal{W}$ on $\mathcal{P}_2(\mathbb{R}^d)$ in the general case is pretty involved (see [2, Definition 10.3.1]) and we shall not need to recall it here. In the next subsection, we follow the approach of [2] to characterize the (unique) element of the *minimal subdifferential of* $\mathcal{W}$ denoted by $\partial^0\mathcal{W}(\mu)$.

2.1. **Subdifferential of $\mathcal{W}$.** Given W a potential satisfying **(NL0)-(NL3)**, let $\partial W(x)$ be the (possibly multivalued) subdifferential of W at the point x, namely the convex set

$$\partial W(x) := \left\{ \kappa \in \mathbb{R}^d : \ W(y) - W(x) \ge \kappa \cdot (y-x) + o(|x-y|), \text{ for all } y \in \mathbb{R}^d \right\}.$$

Denoting by $\partial^0 W(x)$ the (unique) element of $\partial W(x)$ with minimal norm, due to the assumptions **(NL0)-(NL1)** and **(NL3)** we have $\partial^0 W(x) = \nabla W(x)$ for $x \ne 0$ and $\partial^0 W(0) = 0$.

A vector field $\mathbf{w} \in L^2(\mu)$ is said to be an element of the subdifferential of $\mathcal{W}$ at μ, and we write $\mathbf{w} \in \partial\mathcal{W}[\mu]$, if

$$\mathcal{W}[\nu] - \mathcal{W}[\mu] \ge \inf_{\gamma_o \in \Gamma_o(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{w}(x) \cdot (y-x)\, d\gamma_o(x,y) + o(d_W(\nu,\mu)). \tag{2.9}$$

In principle, according to [2, Definition 10.3.1], the elements of $\partial\mathcal{W}[\mu]$ are *plans* γ in the set $\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ such that $(\pi_1)_\#\gamma = \mu$. If a plan $\gamma \in \partial\mathcal{W}[\mu]$ is concentrated on the graph of a vector field $\mathbf{w} \in L^2(\mu)$, then [2, Definition 10.3.1] reduces to (2.9). By following the approach of [2, Sections 10.3 and 10.4], it is easy to see that the (unique) element with minimal norm of $\partial\mathcal{W}[\mu]$ is concentrated on the graph of a vector field. Following [2], we call this element the *minimal subdifferential* of $\mathcal{W}$ at μ, and we denote it by $\partial^0\mathcal{W}[\mu]$. The following characterization of the subdifferential is obtained in [2, Theorem 10.4.11] for smooth C^1-potentials, and here we generalize it to potentials satisfying **(NL0)-(NL3)**:

Proposition 2.2. *Given a potential satisfying* **(NL0)-(NL3)***, the vector field*

$$\kappa(x) := (\partial^0 W * \mu)(x) = \int_{y \ne x} \nabla W(x-y)\, d\mu(y)$$

is the unique element of minimal $L^2(\mu)$-norm in the subdifferential of $\mathcal{W}$, i.e. $\partial^0 W * \mu = \partial^0 \mathcal{W}[\mu]$.

Proof. We divide the proof into two steps.
Step 1: $\kappa(x) \in \partial\mathcal{W}[\mu]$. We have to show that

$$\mathcal{W}[\nu] - \mathcal{W}[\mu] \geq \inf_{\gamma \in \Gamma_o(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \kappa(x) \cdot (y-x)\, d\gamma(x,y) + o(d_W(\nu,\mu)).$$

Thanks to the λ-convexity of W, it suffices to prove that, for any fixed $\gamma \in \Gamma_o(\mu,\nu)$, we have

$$\liminf_{t\to 0} \frac{\mathcal{W}[(1-t)\pi_1 + t\pi_2)_{\#}\gamma] - \mathcal{W}[\mu]}{t} \geq \int_{\mathbb{R}^d \times \mathbb{R}^d} \kappa(x) \cdot (y-x)\, d\gamma(x,y). \tag{2.10}$$

To see this, we observe that the λ–convexity of W implies that the function

$$t \mapsto f(t) := \frac{W(ty + (1-t)x) - W(x)}{t} - \frac{\lambda}{2} t|x-y|^2 \tag{2.11}$$

is nondecreasing in t. Therefore, by writing $f(1) \geq \liminf_{t\searrow 0} f(t)$, integrating with respecto to γ, and using the monotone convergence theorem, we easily recover

$$\mathcal{W}[\nu] - \mathcal{W}[\mu] \geq \liminf_{t\to 0} \frac{\mathcal{W}[(1-t)\pi_1 + t\pi_2)_{\#}\gamma] - \mathcal{W}[\mu]}{t} + \frac{\lambda}{2} d_W^2(\nu,\mu).$$

We now prove (2.10). Let us write $W = \tilde{W} + \frac{\lambda}{2}|x|^2$, so that $\tilde{W} := W - \frac{\lambda}{2}|x|^2$ is convex and $0 \in \partial\tilde{W}(0)$. Moreover we define

$$\tilde{\mathcal{W}}[\mu] := \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \tilde{W}(x-y) d\mu(x) d\mu(y), \qquad \mathcal{Q}[\mu] := \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\mu(x) d\mu(y).$$

Observe that $\mathcal{W} = \tilde{\mathcal{W}} + \lambda\mathcal{Q}$. We first estimate $\frac{1}{t}\big(\tilde{\mathcal{W}}[(1-t)\pi_1 + t\pi_2)_{\#}\gamma] - \tilde{\mathcal{W}}[\mu]\big)$: since $\tilde{W}$ is nonnegative, we have

$$\begin{aligned}
&\frac{\tilde{\mathcal{W}}[(1-t)\pi_1 + t\pi_2)_{\#}\gamma] - \tilde{\mathcal{W}}[\mu]}{t} \\
&= \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} \int_{\mathbb{R}^d \times \mathbb{R}^d} \left[\frac{\tilde{W}(t(y_2 - y_1) + (1-t)(x_2 - x_1)) - \tilde{W}(x_2 - x_1)}{t} \right] d\gamma(x_1,y_1)\, d\gamma(x_2,y_2) \\
&\geq \frac{1}{2} \int_{x_1 \neq x_2} \left[\frac{\tilde{W}(t(y_2 - y_1) + (1-t)(x_2 - x_1)) - \tilde{W}(x_2 - x_1)}{t} \right] d\gamma(x_1,y_1)\, d\gamma(x_2,y_2).
\end{aligned}$$

Thanks to the convexity of $\tilde{W}$ and its (at most) quadratic growth at infinity, and using the fact that $\nabla\tilde{W}$ is odd, it is easily seen that the last term in the above equation converges to

$$\int_{x_1 \neq x_2} \nabla\tilde{W}(x_2 - x_1) \cdot (y_2 - x_2)\, d\gamma(x_1,y_1)\, d\gamma(x_2,y_2).$$

On the other hand, it is an easy computation to check that

$$\begin{aligned}
\frac{\mathcal{Q}[(1-t)\pi_1 + t\pi_2)_{\#}\gamma] - \mathcal{Q}[\mu]}{t} &\to \int_{\mathbb{R}^d \times \mathbb{R}^d} (x_2 - x_1) \cdot (y_2 - x_2)\, d\gamma(x_1,y_1)\, d\gamma(x_2,y_2) \\
&= \int_{x_1 \neq x_2} (x_2 - x_1) \cdot (y_2 - x_2)\, d\gamma(x_1,y_1)\, d\gamma(x_2,y_2).
\end{aligned}$$

Combining all these estimates together, we get the desired result.

Step 2: w is the element of minimal norm of $\partial\mathcal{W}[\mu]$. We closely follows the argument in [2, Theorem 10.4.11]. Fix a vector field $\xi \in C_c^\infty(\mathbb{R}^d, \mathbb{R}^d)$. Observing that $W(x - z + t(\xi(x) - \xi(z))) = W(x - z) = 0$ when $x = z$, we get

$$\begin{aligned}
&\lim_{t\to 0} \frac{\mathcal{W}[(id + t\xi)_\#\mu] - \mathcal{W}[\mu]}{t} \\
&= \lim_{t\to 0} \frac{1}{2} \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{W((x - z) + t(\xi(x) - \xi(z))) - W(x - z)}{t} \, d\mu(x)\, d\mu(z) \\
&= \lim_{t\to 0} \frac{1}{2} \int_{x\neq z} \frac{W((x - z) + t(\xi(x) - \xi(z))) - W(x - z)}{t} \, d\mu(x)\, d\mu(z) \\
&= \frac{1}{2} \int_{x\neq z} \nabla W(x - z) \cdot (\xi(x) - \xi(z))\, d\mu(x)\, d\mu(z) \\
&= \int_{\mathbb{R}^d} \kappa(x) \cdot \xi(x)\, d\mu(x).
\end{aligned} \tag{2.12}$$

Hence, since the definition of slope (2.4) easily implies

$$\liminf_{t\searrow 0} \frac{\mathcal{W}[(id + t\xi)_\#\mu] - \mathcal{W}[\mu]}{d_W((id + t\xi)_\#\mu, \mu)} \geq -|\partial\mathcal{W}|(\mu),$$

we can use (2.8) and (2.12) to get

$$\int_{\mathbb{R}^d} \kappa(x) \cdot \xi(x)\, d\mu(x) \geq -|\partial\mathcal{W}|(\mu) \liminf_{t\to 0} \frac{d_W((id + t\xi)_\#\mu, \mu)}{t} \geq -|\partial\mathcal{W}|(\mu)\|\xi\|_{L^2(\mu)}.$$

Changing ξ with $-\xi$ gives

$$\left| \int_{\mathbb{R}^d} \kappa(x) \cdot \xi(x)\, d\mu(x) \right| \leq |\partial\mathcal{W}|(\mu)\|\xi\|_{L^2(\mu)},$$

so that by the arbitrariness of ξ we get $\|\kappa\|_{L^2(\mu)} \leq |\partial\mathcal{W}|(\mu)$, and therefore κ is the (unique) element of minimal norm. □

2.2. **Well-posedness and convergence of the scheme.** The approach of [2] in proving the existence of a curve of maximal slope for a functional on $\mathcal{P}_2$ is based on a variational version of the implicit Euler scheme, sometimes referred to as the *Jordan–Kinderlehrer–Otto (JKO) scheme* or *minimizing movement scheme* [25, 1, 2]. Given an initial measure $\mu_0 \in \mathcal{P}_2$ and time-step $\tau > 0$, we consider a sequence μ_k^τ recursively defined by $\mu_0^\tau = \mu_0$ and

$$\mu_{k+1}^\tau \in \arg\min_{\mu\in\mathcal{P}_2} \left\{ \mathcal{W}[\mu] + \frac{1}{2\,\tau} d_W^2(\mu_k^\tau, \mu) \right\} , \tag{2.13}$$

for all $k \in \mathbb{N}$.

We shall address here the well-posedness of the definition (2.13) and the convergence of μ_k^τ as $\tau \to 0$ (after a suitable interpolation) to a limit which satisfies Definition 2.1. Such a problem has been widely studied for smooth convex potentials in [2], where convergence of the discrete scheme to a suitable limit is shown. However, allowing for $W(x)$ behaving like $-C|x|^2$ as $|x| \to +\infty$ and for a pointy singularity at $x = 0$ would require in general some improvements of the arguments in [2, Part I], as we shall see below. Indeed let us point out that, for $W(x)$ behaving like $-C|x|^2$, the functional $\mathcal{W}[\mu]$ is upper (and not lower!) semicontinuous with respect to the narrow convergence. Let us observe that in our case one could exploit the fact the functional $\mathcal{W}$ is λ-convex along generalized geodesic to directly apply the theory developed in [2], see Remark 2.9. On the other hand our proofs are more flexible, and so we believe that our strategy can be of interest also in other situations.

For the sake of clarity, we shall recall all the main steps of the JKO scheme developed in [2] in the particular case of a functional given by a pure nonlocal interaction energy. We shall perform this task also for another reason, namely to relax the set of assumptions **(NL0)-(NL3)** in order to admit $|\nabla W|$ to be possibly unbounded at the origin (see Remark 2.11).

We start by showing that the minimization problem (2.13) admits at least one solution, which in our situation is not a trivial issue. To this aim, we prove a technical lemma which will be also useful in the sequel.

Lemma 2.3 (Weak lower semi–continuity of the penalized interaction energy)**.** *Suppose W satisfies* **(NL0)-(NL3)***. Then, for a fixed $\overline{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$, the penalized interaction energy functional*

$$\mathcal{P}_2(\mathbb{R}^d) \ni \mu \mapsto \mathcal{W}[\mu] + \frac{1}{2\tau} d_W^2(\mu, \overline{\mu})$$

is lower semi–continuous with respect to the narrow topology of $\mathcal{P}(\mathbb{R}^d)$ for all $\tau > 0$ such that $8\tau\lambda^- \leq 1$, where $\lambda^- := \max\{0, -\lambda\}$.

Proof. Let $\{\mu_n\}_n \subset \mathcal{P}_2(\mathbb{R}^d)$ such that $\lim_{n\to+\infty} \mu_n = \mu_\infty$ narrowly. We have to prove that

$$\liminf_{n\to+\infty} \left[\mathcal{W}[\mu_n] + \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) \right] \geq \mathcal{W}[\mu_\infty] + \frac{1}{2\,\tau} d_W^2(\mu_\infty, \overline{\mu}). \tag{2.14}$$

From the estimate in Remark 1.1 and the fact that $\lambda \leq 0$, we have

$$W(x-y) \geq \frac{\lambda}{2}|x-y|^2 \geq \lambda(|x|^2 + |y|^2),$$

which implies that

$$h(x,y) := W(x-y) - \lambda(|x|^2 + |y|^2)$$

is a nonnegative continuous function. Therefore,

$$\begin{aligned} \mathcal{W}[\mu_n] + \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) &= \lambda \int_{\mathbb{R}^d\times\mathbb{R}^d} (|x|^2 + |y|^2)\, d\mu_n(x)\, d\mu_n(y) \\ &\quad + \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x,y)\, d\mu_n(x)\, d\mu_n(y) + \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x,y)\, d\mu_n(x)\, d\mu_n(y) \\ &\quad + \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) + 2\lambda \int_{\mathbb{R}^d} |x|^2\, d\mu_n(x). \end{aligned} \tag{2.15}$$

Since $h \geq 0$, we easily get

$$\liminf_{n\to+\infty} \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x,y)\, d\mu_n(x)\, d\mu_n(y) \geq \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x,y)\, d\mu_\infty(x)\, d\mu_\infty(y). \tag{2.16}$$

Therefore, to get the desired assertion it suffices to prove that

$$\liminf_{n\to+\infty} \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) + 2\lambda \int_{\mathbb{R}^d} |x|^2\, d\mu_n(x) \geq \frac{1}{2\,\tau} d_W^2(\mu_\infty, \overline{\mu}) + 2\lambda \int_{\mathbb{R}^d} |x|^2\, d\mu_\infty(x). \tag{2.17}$$

Now, let $\gamma_n \in \Gamma_o(\overline{\mu}, \mu_n)$. Then,

$$\frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) + 2\lambda \int_{\mathbb{R}^d} |x|^2 d\mu_n(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{1}{2\tau}|x-y|^2 + 2\lambda |y|^2 \right) d\gamma_n(x,y). \tag{2.18}$$

Stability of optimal transportation plans (see [45, Theorem 5.20]) implies that there exists a subsequence, that we may assume to be the whole sequence, such that γ_n converges narrowly to an optimal plan $\gamma_\infty \in \Gamma_o(\overline{\mu}, \mu_\infty)$. As a consequence of

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} |x|^2 \, d\gamma_n(x,y) = \int_{\mathbb{R}^d} |x|^2 \, d\overline{\mu}(x) = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x|^2 \, d\gamma_\infty(x,y)$$

and of the elementary inequality $|y|^2 \le 2|x-y|^2 + 2|x|^2$ which implies

$$\frac{1}{2\tau}|x-y|^2 + 2\lambda|y|^2 + \frac{1}{2\tau}|x|^2 \ge \left(\frac{1}{4\tau} + 2\lambda \right) |y|^2 \ge 0 \qquad \text{if} \;\; \tau \le \frac{1}{8\lambda^-}, \tag{2.19}$$

we easily obtain

$$\begin{aligned}
&\liminf_{n\to+\infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{1}{2\tau}|x-y|^2 + 2\lambda|y|^2 \right) d\gamma_n(x,y) \\
&= -\frac{1}{2\tau} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x|^2 d\gamma_\infty(x,y) + \liminf_{n\to+\infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{1}{2\tau}|x-y|^2 + 2\lambda|y|^2 + \frac{1}{2\tau}|x|^2 \right) d\gamma_n(x,y) \\
&\ge -\frac{1}{2\tau} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x|^2 d\gamma_\infty(x,y) + \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{1}{2\tau}|x-y|^2 + 2\lambda|y|^2 + \frac{1}{2\tau}|x|^2 \right) d\gamma_\infty(x,y) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(\frac{1}{2\tau}|x-y|^2 + 2\lambda|y|^2 \right) d\gamma_\infty(x,y).
\end{aligned}$$

This proves (2.17). □

Remark 2.4. *We observe that the optimality of the plans γ_n and γ_∞ is never actually needed in the previous proof. More precisely, the weak lower semi–continuity property stated in the above lemma still holds for the functional*

$$\mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \ni \gamma \mapsto \mathcal{W}[(\pi_1)_\# \gamma] + \frac{1}{2\tau} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 \, d\gamma(x,y),$$

where $(\pi_2)_\# \gamma = \overline{\mu}$ is fixed.

Next, we prove the solvability of the minimization problem (2.13).

Proposition 2.5 (Existence of minimizers)**.** *Suppose W satisfies* **(NL0)-(NL3)***. Then, there exists $\tau_0 > 0$ depending only on W such that, for all $0 < \tau < \tau_0$ and for a given $\overline{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$, there is $\mu_\infty \in \mathcal{P}_2(\mathbb{R}^d)$ such that*

$$\mathcal{W}[\mu_\infty] + \frac{1}{2\,\tau} d_W^2(\overline{\mu}, \mu_\infty) = \min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \mathcal{W}[\mu] + \frac{1}{2\,\tau} d_W^2(\overline{\mu}, \mu) \right\}.$$

Proof. Step 1: Compactness. Let us fix a measure $\overline{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ and a time step $\tau > 0$, and consider a minimizing sequence $\mu_n \in \mathcal{P}_2(\mathbb{R}^d)$, i.e.

$$\inf_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \left\{ \mathcal{W}[\mu] + \frac{1}{2\,\tau} d_W^2(\mu, \overline{\mu}) \right\} = \lim_{n\to+\infty} \left\{ \mathcal{W}[\mu_n] + \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) \right\}.$$

Since μ_n is a minimizing sequence, we have

$$\mathcal{W}[\mu_n] + \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu}) \leq C_1 \tag{2.20}$$

for some constant C_1. Then, the lower estimate of W in Remark 1.1 and the inequality (2.19) imply, after very similar computations as in Lemma 2.3, that $d_W^2(\mu_n, \overline{\mu})$ is uniformly bounded with respect to n if τ is small enough. Prokhorov's compactness theorem then implies that the sequence $\{\mu_n\}_n$ is tight.

STEP 2: COERCIVITY. We need to prove that

$$\liminf_{n\to+\infty} \left[\mathcal{W}[\mu_n] + \frac{1}{2\,\tau} d_W^2(\mu_n, \overline{\mu})\right] \geq C_0\, d_W^2(\mu_n, \overline{\mu}) - C_1$$

for some positive constant C_0, C_1 independent on n. This follows similarly to Step 1 for τ small enough.

STEP 3: PASSING TO THE LIMIT BY LOWER SEMI–CONTINUITY. This is a consequence of Lemma 2.3. □

Next we have to establish that the family $\{\mu_k^\tau\}_{\tau\in(0,\tau_0)}$ (up to a suitable interpolation) converges narrowly to a certain limit. This task can be performed exactly as described in [2, Chapters 2, 3]. For the sake of clarity, we recall here the result in [2] stating the convergence of the JKO scheme. The proof can be found in [2, Proposition 2.2.3]. First, we introduce the piecewise constant interpolation

$$\begin{aligned}&\mu^\tau(0) := \mu_0\\ &\mu^\tau(t) := \mu_k^\tau \qquad \text{if} \quad t\in((k-1)\tau, k\tau], \quad k\geq 1.\end{aligned}$$

Proposition 2.6 (Compactness in the JKO scheme [2]). *Suppose W satisfies* **(NL0)-(NL3)**. *There exist a sequence $\tau_n \searrow 0$, and a limit curve $\mu \in AC_{loc}\left([0,+\infty); \mathcal{P}_2(\mathbb{R}^d)\right)$, such that*

$$\mu^n(t) := \mu^{\tau_n}(t) \to \mu(t), \qquad \textit{narrowly as} \quad n\to+\infty$$

for all $t\in[0,+\infty)$.

According to the notation recalled in [2, Definition 2.0.6], the above proposition states that the set of minimizing movements for $\mathcal{W}$ starting from μ_0 is not empty. The last step of the procedure proposed in [2] is to check that the limit curve provided by Proposition 2.6 is a curve of maximal slope for W according to definition 2.1.

Let $\tilde{\mu}^n(t)$ denote the *De Giorgi variational interpolation* (see [2, Section 3.2]). Then, from [2, Equation (3.1.12)] and the argument in the proof of [2, Lemma 3.2.2] we have the energy inequality

$$\mathcal{W}[\mu_0] \geq \frac{1}{2}\int_0^T \|v^n(t)\|^2_{L^2(\mu^n(t))}\,dt + \frac{1}{2}\int_0^T |\partial\mathcal{W}|(\tilde{\mu}^n(t))^2\,dt + \mathcal{W}[\mu^n(T)] \tag{2.21}$$

for all $T>0$, where on any interval $[(k-1)\tau_n, k\tau_n]$ the curve $\mu^n(t)$ is a Wasserstein geodesic connecting $\mu_{k-1}^{\tau_n}$ to $\mu_k^{\tau_n}$, and $v^n(t)$ is its velocity field. Let us recall that the continuity equation $\partial_t\mu^n(t) + \mathrm{div}(v^n(t)\mu^n(t)) = 0$ holds, and that up to a subsequence both $\mu^n(t)$ and $\tilde{\mu}^n(t)$ narrowly converge to the same limit curve $\mu(t)$ on $[0,+\infty)$ provided by Proposition 2.6. The following lemma is needed to suitably pass to the limit the slope term in (2.21).

Lemma 2.7 (Lower semicontinuity of the slope)**.**

$$\liminf_{n\to+\infty}\int_0^T |\partial\mathcal{W}|^2(\tilde\mu^n(t))dt \geq \int_0^T |\partial\mathcal{W}|^2(\mu(t))dt.$$

Proof. By using the representation formula proven in Proposition 2.2, we have to prove that

$$\liminf_{n\to+\infty}\int_0^T\int_{\mathbb{R}^d} |\kappa^n(x,t)|^2 d\tilde\mu^n(t)(x)dt \geq \int_0^T\int_{\mathbb{R}^d} |\kappa(x,t)|^2 d\mu(t)(x)dt,$$

where

$$\kappa^n(x,t) := \partial^0 W * \tilde\mu^n(x,t), \qquad \kappa(x,t) := \partial^0 W * \mu(x,t).$$

Without loss of generality, up to passing to a subsequence we can assume that

$$\sup_n \int_0^T\int_{\mathbb{R}^d} |\kappa^n(x,t)|^2\, d\tilde\mu^n(t)(x)\, dt < +\infty.$$

Hence, as a byproduct of [2, Theorem 5.4.4] on the measure space $X := \mathbb{R}^d \times [0,T]$ with the family of measures $\mu^n \otimes dt$, we get the desired assertion once we prove that κ^n converges *weakly* to κ, i.e. that for any vector field $\phi \in C_c^\infty(\mathbb{R}^d\times[0,T];\mathbb{R}^d)$

$$\int_0^T\int_{\mathbb{R}^d} \phi(x,t)\cdot\kappa^n(x,t)d\tilde\mu^n(t)(x) \to \int_0^T\int_{\mathbb{R}^d} \phi(x,t)\cdot\kappa(x,t)d\mu(t)(x) \tag{2.22}$$

as $n\to+\infty$. To show this, we observe that term on the left-hand side is given by

$$\begin{aligned}\int_0^T\int_{\mathbb{R}^d} \phi(x,t)\cdot\kappa^n(x,t)\, d\tilde\mu^n(t)(x) &= \int_0^T\int_{x\neq y} \phi(x,t)\cdot\nabla W(x-y)\, d\tilde\mu^n(t)(y)\, d\tilde\mu^n(t)(x)\, dt\\ &= \frac{1}{2}\int_0^T\int_{x\neq y} (\phi(x,t)-\phi(y,t))\cdot\nabla W(x-y)\, d\tilde\mu^n(t)(y)\, d\tilde\mu^n(t)(x)\, dt,\end{aligned}$$

where for the second equality we used the (crucial) fact that ∇W is odd, so we could symmetrize the expression inside the integral.

By [2, Lemma 3.2.2], the sequence μ_n has uniformly bounded second moments. Therefore, thanks to the linear growth control on the gradient of W in (1.5), the function $(\phi(x,t)-\phi(y,t))\cdot\nabla W(x-y)$ is uniformly integrable with respect to $\tilde\mu^n(t)\otimes\tilde\mu^n(t)\otimes dt$, and we easily recover (2.22) by weak convergence arguments. □

We are now ready to complete the proof of the existence of a solution to (1.2)–(1.3) in the sense of Definition 2.1.

Theorem 2.8 (Existence of curves of maximal slope)**.** *Let W satisfy the assumptions* **(NL0)-(NL3)***. Then, there exists at least one curve of maximal slope for the functional $\mathcal{W}$, i.e. there exists at least one curve $\mu \in AC_{loc}\left([0,+\infty);\mathcal{P}_2(\mathbb{R}^d)\right)$ such that the energy inequality*

$$\begin{aligned}\mathcal{W}[\mu_0] \geq \frac{1}{2}\int_0^T \|v(t)\|^2_{L^2(\mu(t))}\, dt\\ + \frac{1}{2}\int_0^T\int_{\mathbb{R}^d}\left|\int_{x\neq y}\nabla W(x-y)d\mu(t)(y)\right|^2 d\mu(t)(x)\, dt + \mathcal{W}[\mu(T)],\end{aligned} \tag{2.23}$$

is satisfied, where $v(t)\in L^2(\mu(t))$ is the minimal velocity field associated to μ.

Proof. We want to prove that the curve $\mu(t)$ provided by Proposition 2.6 satisfies the desired condition. As a consequence of (2.21) and of Lemma 2.7, if we show that

$$\liminf_{n\to\infty} \frac{1}{2}\int_0^T \|v^n(t)\|^2_{L^2(\mu^n(t))}\,dt + \mathcal{W}[\mu^n(T)] \geq \frac{1}{2}\int_0^T \|v(t)\|^2_{L^2(\mu^n(t))}\,dt + \mathcal{W}[\mu(T)], \quad (2.24)$$

all the remaining parts of the proof of the convergence of the scheme to a solution goes through like in the case when $\mathcal{W}$ is lower semicontinuous with respect to the narrow topology, see [2, Chapter 3].

To prove the inequality (2.24), having in mind the constitutive relation (2.7) linking μ^n and v^n, we regularize the solutions of $\partial_t\mu^n(t) + \operatorname{div}(v^n(t)\mu^n(t)) = 0$ and $\partial_t\mu(t) + \operatorname{div}(v(t)\mu(t)) = 0$ as follows:

$$\begin{aligned} v^{n,\varepsilon}(t) &:= \frac{(v^n(t)\mu^n(t)) * \eta_\varepsilon}{\mu^n(t) * \eta_\varepsilon}, \qquad & \mu^{n,\varepsilon}(t) &:= \mu^n(t) * \eta_\varepsilon,\\ v^{\varepsilon}(t) &:= \frac{(v(t)\mu(t)) * \eta_\varepsilon}{\mu(t) * \eta_\varepsilon}, \qquad & \mu^{\varepsilon}(t) &:= \mu(t) * \eta_\varepsilon, \end{aligned}$$

where $\eta_\varepsilon = \frac{1}{\varepsilon^d}\eta\left(\frac{\cdot}{\varepsilon}\right) \in C^\infty(\mathbb{R}^d)$ is a smooth convolution kernel with support the whole $\mathbb{R}^d$, say a gaussian. Applying [2, Proposition 8.1.8] we deduce that the measures $\mu^{n,\varepsilon}(t)$, $\mu^\varepsilon(t)$ are given by the formula $\mu^{n,\varepsilon}(t) = (X^{n,\varepsilon}(t))_{\#}\mu_0$ and $\mu^\varepsilon(t) = (X^\varepsilon(t))_{\#}\mu_0$, where $X^{n,\varepsilon}(t)$ and $X^\varepsilon(t)$ denote the flows of $v^{n,\varepsilon}(t)$ and $v^\varepsilon(t)$ respectively, more precisely

$$\begin{aligned} &\frac{d}{dt}X^{n,\varepsilon}(t,x) = v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x)), \qquad X^{n,\varepsilon}(0,x) = x,\\ &\frac{d}{dt}X^{\varepsilon}(t,x) = v^{\varepsilon}(t, X^{\varepsilon}(t,x)), \qquad X^{\varepsilon}(0,x) = x. \end{aligned}$$

We now define the transport map from $\mu^\varepsilon(T)$ to $\mu^{n,\varepsilon}(T)$ as $T_n^\varepsilon := X^{n,\varepsilon}(T) \circ (X^\varepsilon(T))^{-1}$. We have

$$\begin{aligned} &d_W^2(\mu^\varepsilon(T), \mu^{n,\varepsilon}(T)) \leq \int_{\mathbb{R}^d} |T_n^\varepsilon(x) - x|^2\, d\mu^\varepsilon(T)(x)\\ &= \int_{\mathbb{R}^d} |X^{n,\varepsilon}(T)\circ(X^\varepsilon(T))^{-1}(x) - (X^\varepsilon(T))^{-1}(x) + (X^\varepsilon(T))^{-1}(x) - x|^2\, d\mu(T)(x)\\ &= \int_{\mathbb{R}^d} \left|\int_0^T \big[v^{n,\varepsilon}(t, X^{n,\varepsilon}(t)\circ(X^\varepsilon(T))^{-1}(x)) - v^\varepsilon(t, X^\varepsilon(t)\circ(X^\varepsilon(T))^{-1}(x))\big]\,dt\right|^2 d\mu^\varepsilon(T)(x)\\ &= \int_{\mathbb{R}^d} \left|\int_0^T \big[v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x)) - v^\varepsilon(t, X^\varepsilon(t,x))\big]\,dt\right|^2\, d\mu_0^\varepsilon(x) \end{aligned}$$

By Hölder's inequality and expanding the squares, we get

$$\begin{aligned} d_W^2(\mu^\varepsilon(T), \mu^{n,\varepsilon}(T)) &\leq T\int_{\mathbb{R}^d}\int_0^T \left|v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x)) - v^\varepsilon(t, X^\varepsilon(t,x))\right|^2 dt\, d\mu_0^\varepsilon(x)\\ &\leq T\int_0^T\int_{\mathbb{R}^d} |v^{n,\varepsilon}(t,x)|^2\, d\mu^{n,\varepsilon}(t)(x)\,dt + T\int_0^T\int_{\mathbb{R}^d} |v^\varepsilon(t,x)|^2\, d\mu^\varepsilon(t)(x)\,dt\\ &\quad - 2T\int_0^T\int_{\mathbb{R}^d} v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x))\cdot v^\varepsilon(t, X^\varepsilon(t,x))\, d\mu_0(x)\,dt. \qquad (2.25) \end{aligned}$$

Thanks to [2, Lemma 8.1.10] we have

$$\int_0^T \int_{\mathbb{R}^d} |v^{n,\varepsilon}(t,x)|^2 \, d\mu^{n,\varepsilon}(t)(x)\, dt \le \int_0^T \int_{\mathbb{R}^d} |v^n(t,x)|^2 \, d\mu^n(t)(x)\, dt \qquad \forall\, \varepsilon > 0. \tag{2.26}$$

Moreover, thanks to the weak convergence of $(\mu^n(t), v^n(t)\mu^n(t))$ to $(\mu(t), v(t)\mu(t))$, which is a consequence of the linear growth control of the gradient of W in (1.5) and the fact that $\mu^{n,\varepsilon}(t)$ and $\mu^\varepsilon(t)$ are uniformly (in $n \in \mathbb{N}$) bounded away from zero on compact sets of $\mathbb{R}^d$, we deduce that

$$v^{n,\varepsilon}(t) \to v^\varepsilon(t) \;\text{ in }\; L^1([0,T], C^\infty_{loc}(\mathbb{R}^d)). \tag{2.27}$$

Indeed,

$$\begin{aligned} D^\alpha[v^{n,\varepsilon} - v^\varepsilon] &= \frac{D^\alpha \eta^\varepsilon * (v^n\mu^n)}{\mu^{n,\varepsilon}} - \frac{D^\alpha \eta^\varepsilon * (v\mu)}{\mu^\varepsilon} \\ &= D^\alpha \eta^\varepsilon * (v^n\mu^n) \left(\frac{\mu^\varepsilon - \mu^{n,\varepsilon}}{\mu^\varepsilon \mu^{n,\varepsilon}} \right) + \frac{1}{\mu^\varepsilon} D^\alpha \eta^\varepsilon * (v\mu - v^n\mu^n) \end{aligned}$$

and v^n is uniformly bounded in $L^2(\mu^n)$ with respect to n. Since the flows $X^{n,\varepsilon}(t)$ and $X^\varepsilon(t)$ are globally defined (see for instance [2, Proposition 8.1.8]), (2.27) easily implies that for any $t \in [0,T]$

$$X^{n,\varepsilon}(t) \to X^\varepsilon(t) \quad \text{locally uniformly on compact subsets of } \mathbb{R}^d. \tag{2.28}$$

This fact, together with the fact that $v^{n,\varepsilon}(t, X^{n,\varepsilon}(t))$ are uniformly bounded in $L^2(\mu_0 \otimes dt)$ thanks to (2.26), implies that

$$\begin{aligned} &\lim_{n\to\infty} \int_{\mathbb{R}^d} \int_0^T v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x)) \cdot v^\varepsilon(t, X^\varepsilon(t,x))\, dt\, d\mu_0(x) \\ &\quad = \int_{\mathbb{R}^d} \int_0^T |v^\varepsilon(t, X^\varepsilon(t,x))|^2\, dt\, d\mu_0(x) = \int_{\mathbb{R}^d} \int_0^T |v^\varepsilon(t,x)|^2\, d\mu_0(x)\, dt. \end{aligned} \tag{2.29}$$

To prove (2.29), split the integral on the left-hand side as follows

$$\begin{aligned} &\int_{\mathbb{R}^d} \int_0^T v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x)) \cdot v^\varepsilon(t, X^\varepsilon(t,x))\, dt\, d\mu_0(x) \\ &\qquad = \int_{|x|>R} \int_0^T v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x)) \cdot v^\varepsilon(t, X^\varepsilon(t,x))\, dt\, d\mu_0(x) \\ &\qquad\quad + \int_{|x|\le R} \int_0^T v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x)) \cdot v^\varepsilon(t, X^\varepsilon(t,x))\, dt\, d\mu_0(x) =: I_1 + I_2. \end{aligned}$$

Now, thanks to (2.26) and the fact that v^n is uniformly bounded in $L^2(\mu^n)$ with respect to n, we can estimate

$$\begin{aligned} I_1^2 &\le \int_{\mathbb{R}^d} \int_0^T |v^{n,\varepsilon}(t, X^{n,\varepsilon}(t,x))|^2\, dt\, d\mu_0(x) \int_{|x|>R} \int_0^T |v^\varepsilon(t, X^\varepsilon(t,x))|^2\, dt\, d\mu_0(x) \\ &\le C \int_{|x|>R} \int_0^T |v^\varepsilon(t, X^\varepsilon(t,x))|^2\, dt\, d\mu_0(x) \end{aligned}$$

for some constant C independent on n. Hence, one can choose R large enough such that $|I_1| < \eta$ for an arbitrarily small $\eta > 0$. On the other hand, (2.27) and (2.28) imply

$$I_2 \to \int_{|x|\le R} \int_0^T |v^\varepsilon(t,x)|^2 \, d\mu_0(x)\, dt$$

as $n \to +\infty$, and (2.29) follows by letting $R \to +\infty$.

Therefore, by combining (2.29) with (2.25) and (2.26) we obtain

$$\begin{aligned}&\liminf_{n\to\infty} d_W^2(\mu^\varepsilon(T), \mu^{n,\varepsilon}(T)) + 2T\mathcal{W}[\mu^n(T)] \\ &\le \liminf_{n\to\infty} T\left[\int_0^T \int_{\mathbb{R}^d} |v^n(t,x)|^2 \, d\mu^n(t)(x)\, dt - \int_0^T \int_{\mathbb{R}^d} |v^\varepsilon(t,x)|^2 \, d\mu^\varepsilon(t)(x)\, dt + 2\mathcal{W}[\mu^n(T)]\right].\end{aligned} \tag{2.30}$$

We now claim that there exists a constant $C_0 > 0$, depending only on the convolution kernel η, such that for any $\mu \in \mathcal{P}(\mathbb{R}^d)$

$$d_W^2(\mu, \mu * \eta_\varepsilon) \le C_0 \varepsilon^2. \tag{2.31}$$

Indeed is suffices to consider the transport plan $\gamma^\varepsilon \in \Gamma(\mu, \mu * \eta_\varepsilon)$ defined has

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} f(x,y)\, d\gamma^\varepsilon(x,y) := \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x,y)\eta_\varepsilon(y-x)\, dy\, d\mu(x) \qquad \forall\, f \in C_b(\mathbb{R}^d\times\mathbb{R}^d),$$

to get that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^2 \, d\gamma^\varepsilon(x,y) = \int_{\mathbb{R}^d} |z|^2 \eta_\varepsilon(z)\, dz = \varepsilon^2 \int_{\mathbb{R}^d} |z|^2 \eta(z)\, dz,$$

which proves (2.31). We finally observe that

$$\liminf_{\varepsilon\to 0} \int_0^T \int_{\mathbb{R}^d} |v^\varepsilon(t,x)|^2 \, d\mu^\varepsilon(t)(x)\, dt \ge \int_0^T \int_{\mathbb{R}^d} |v(t,x)|^2 \, d\mu(t)(x)\, dt \tag{2.32}$$

(actually using (2.26) one could prove that the above liminf is a limit, and equality holds). Combining (2.30) with (2.31) we obtain

$$\begin{aligned}\liminf_{n\to\infty} d_W^2(\mu(T), \mu^n(T)) + 2T\mathcal{W}[\mu^n(T)] \le \liminf_{n\to\infty} T\Bigg[\int_0^T \int_{\mathbb{R}^d} |v^n(t,x)|^2 \, d\mu^n(t)\, dt \\ - \int_0^T \int_{\mathbb{R}^d} |v^\varepsilon(t,x)|^2 \, d\mu^\varepsilon(t)(x)\, dt + 2\mathcal{W}[\mu^n(T)]\Bigg] + O(\varepsilon),\end{aligned}$$

so, that letting $\varepsilon \to 0$, thanks to (2.32) we finally get

$$\begin{aligned}&\liminf_{n\to\infty} d_W^2(\mu(T), \mu^n(T)) + 2T\mathcal{W}[\mu^n(T)] \\ &\le \liminf_{n\to\infty} T\left[2\mathcal{W}[\mu^n(T)] + \int_0^T \int_{\mathbb{R}^d} |v^n(t,x)|^2 \, d\mu^n(t)(x)\, dt - \int_0^T \int_{\mathbb{R}^d} |v(t,x)|^2 \, d\mu(t)(x)\, dt\right].\end{aligned} \tag{2.33}$$

Moreover, in view of Lemma 2.3 we deduce

$$\liminf_{n\to\infty} d_W^2(\mu(T), \mu^n(T)) + 2T\mathcal{W}[\mu^n(T)] \ge 2T\mathcal{W}[\mu(T)] \tag{2.34}$$

for T small enough. Combining (2.34) with (2.33), we obtain that (2.24) holds provided T is sufficiently small (but independent on the initial datum μ_0), and this allows to prove the existence of a curve of maximal slope on a small time interval $[0,T]$. Iterating now the construction via minimizing movements on $[T, 2T]$, $[2T, 3T]$ and so on, and adding the energy inequalities (2.23) on each time interval, we get the desired result. □

Remark 2.9 (λ-Generalised convexity)**.** Let us emphasize that, since $\lambda \le 0$, our functional is not only λ-convex with respect to Wasserstein geodesics, but also with respect to generalized geodesics, see [2, Definition 9.2.4]. It follows directly from [2, Proposition 9.3.5], decomposing $\mathcal{W} = \tilde{\mathcal{W}} + \lambda \mathcal{Q}$ as in Proposition 2.2. Exploiting this fact, the existence of solutions for the discrete scheme and the convergence of the scheme follow from the theory developed in [2]. On the other hand, we believe that our strategy to show the convergence of the discrete scheme is of interest in itself, being much more flexible and since it may be applied to more general situations where λ-convexity fails. This is why we chose to prove existence of curves of maximal slope in this way.

Remark 2.10 (The ODE system)**.** Let $x_i(t)$, $i = 1, \dots, N$, be C^1-solutions of the ODE system (for the time intervals that such exist)

$$\dot{x}_i = -\sum_{j \neq i} m_j \nabla W(x_i - x_j), \qquad i = 1, \dots, N, \tag{2.35}$$

with $m_i > 0$ and $\sum_i m_i = 1$. Then it is straightforward to check that $\mu(t) := \sum_{i=1}^N m_i \delta_{x_i(t)}$ is a solution of (1.2) in the sense of Definition 1.2. Conversely, if $\mu(t)$ of the form above solves the PDE and $x_i(t)$ are C^1 curves for $i = 1, \dots, N$, then $x_i(t)$ solve the ODE system.

The question is what happens if the particles collide: can the solutions of the PDE still be represented by an ODE? This question has a positive answer, see for instance [2, Theorems 8.2.1 and 11.2.3 and Equation (11.2.22)] For completeness, we give a sketch a proof in our particular case.

We consider absolutely continuous solutions of

$$\dot{x}_i = -\sum_{j \in C(i)} m_j \nabla W(x_i - x_j), \qquad i = 1, \dots, N, \tag{2.36}$$

$$\text{with } C(i) := \{j \in \{1, \dots, N\} \,:\, j \neq i,\, x_j(t) \neq x_i(t)\}. \tag{2.37}$$

More precisely we consider the solutions of the associated integral equation. If $C(i)$ is empty, then all particles have collapsed to a single particle. We then define the right hand side to be zero, that is we define the sum over empty set of indexes to be zero. The right hand side of this ODE system is bounded and Lipschitz-continuous in space on short time intervals. Thus the ODE system has a unique Lipschitz-continuous solutions on short time intervals. The estimate (1.5) then implies that the solutions are global-in-time. Note that the solutions are Lipschitz (in time) on bounded time intervals. Also note that collisions of particles can occur, but that we do not relabel the particles when they collide. Since the number of particles is N there exist $0 \le k \le N-1$ times $0 =: T_0 < T_1 < T_2 < \cdots < T_k < \infty =: T_{k+1}$ at which collisions occur. Note that $\mu(t) = \sum_{i=1}^N m_i \delta_{x_i(t)}$ is a solution of the PDE on the time intervals $[T_l, T_{l+1})$. Furthermore, the Lipschitz continuity of x_i implies that μ is an absolutely continuous curve in $\mathcal{P}_2(\mathbb{R}^d)$. It is then straightforward to verify that μ is a weak solution according to Definition 1.2. Since the solution to the PDE is unique the converse claim also holds.

Let us mention that while above we did not relabel the particles after collisions, at times it is useful to do so. That is on time intervals $[T_l, T_{l+1})$ the ODE system (2.36) is equivalent to

$$\frac{d\tilde{x}_i}{dt} = -\sum_{j \neq i} \tilde{m}_j \nabla W(\tilde{x}_l - \tilde{x}_j), \qquad i = 1, \dots, N_l, \tag{2.38}$$

where N_l is the number of distinct particles on the time interval $[T_l, T_{l+1})$, and $\tilde{x}_j$, $\tilde{m}_j$ are their locations and masses, respectively.

Remark 2.11 (Existence of minimizing movements when ∇W is unbounded)**.** We remark here that the construction of the JKO scheme, up to the proof of the Proposition 2.6, can be performed even in case ∇W has a singular behavior such as $W(x) = |x|^\alpha$ for $\alpha \in (0,1)$ (although in this case we are not able to characterize the subdifferential). Therefore, one can easily prove that there exist at least one minimizing movement for such a kind of functional. Note that the case $\alpha = 0$ is critical, since one recovers the logarithmic kernel $W(x) = \log|x|$ as $\alpha \to 0$, for which it is an open problem how to define unique global-in-time weak measure solutions for all initial masses, see [39].

2.3. **Gradient Flow Solutions.** In this subsection, we will show the existence of global-in-time weak measure solutions for (1.2) for potentials satisfying **(NL0)-(NL3)** as a consequence of the general abstract theorems proved in [2]. In fact, using that the potential is λ-convex by **(NL1)**, Lemma 2.3 and the existence of minimizers in Proposition 2.5, we meet the hypotheses of [2, Theorem 11.1.3]. This abstract theorem shows that curves of maximal slope are equivalent under certain hypotheses to gradient flows. As a direct consequence of the existence of curves of maximal slope in Theorem 2.8, we can assert the following result. Let us remark that Proposition 2.2 has played a key role in the argument leading to Theorem 2.8 in two ways: allowing to show the lower semicontinuity of the slope to get the energy inequalities, and in order to identify the limiting velocity field.

Theorem 2.12 (Existence of the Gradient Flow)**.** *Let W satisfy the assumptions* **(NL0)-(NL3)**. *Given any $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, then there exists a gradient flow solution, i.e. a curve $\mu \in AC_{loc}([0,\infty); \mathcal{P}_2(\mathbb{R}^d))$ satisfying*

$$\begin{aligned}
&\frac{\partial \mu(t)}{\partial t} + \mathrm{div}(v(t)\mu(t)) = 0 \ \textit{in}\ \mathcal{D}'([0,\infty) \times \mathbb{R}^d),\\
&v(t) = -\partial^0 \mathcal{W}[\mu(t)] = -\partial^0 W * \mu(t),\\
&\|v(t)\|_{L^2(\mu(t))} = |\mu'|(t) \quad \textit{a.e. } t > 0,
\end{aligned}$$

with $\mu(0) = \mu_0$. Moreover, the energy identity

$$\int_a^b \int_{\mathbb{R}^d} |v(t,x)|^2 \, d\mu(t)(x)\, dt + \mathcal{W}[\mu(b)] = \mathcal{W}[\mu(a)] \tag{2.39}$$

holds for all $0 \le a \le b < \infty$.

To summarize, the notions of curves of maximal slope and gradient flow solutions are equivalent and they imply the notion of weak measure solutions in the sense of Definition 1.2. Furthermore, absolute continuity of the curve of weak measure solutions and the characterization of the subdifferential imply that weak measure solutions are gradient flow solutions, see [2, Sections 8.3 and 8.4]. Thus, the three notions of solution are equivalent.

The λ-geodesic convexity of the functional plays a crucial role for the uniqueness of gradient flow solutions. Since the interaction potential is λ-geodesically convex for $\lambda \le 0$, the following result follows readily from [2, Theorem 11.1.4].

Theorem 2.13 (d_W-Stability)**.** *Let W satisfy the assumptions* **(NL0)-(NL3)***. Given two gradient flow solutions $\mu^1(t)$ and $\mu^2(t)$ in the sense of the theorem above, we have*

$$d_W(\mu^1(t),\mu^2(t)) \le e^{-\lambda t}\, d_W(\mu^1_0,\mu^2_0)$$

for all $t \ge 0$. In particular, the gradient flow solution starting from any given $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$ is unique. Moreover, this solution is characterized by a system of evolution variational inequalities:

$$\frac{1}{2}\frac{d}{dt}\, d_W^2(\mu(t),\sigma) + \frac{\lambda}{2}\, d_W^2(\mu(t),\sigma) \le \mathcal{W}[\sigma] - \mathcal{W}[\mu(t)] \quad a.e.\ t>0,$$

for all $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$.

With this we have completed the existence, uniqueness and stability for gradient flow solutions for potentials satisfying **(NL0)-(NL3)**.

Remark 2.14 (Case $\lambda > 0$)**.** All the theory and results obtained in this section can be applied to λ-convex potentials with $\lambda > 0$, i.e. allowing for $\lambda > 0$ in **(NL1)**, provided we restrict ourselves to measures with equal initial center of mass. This relies on the fact that, when $\lambda > 0$, the interaction potential is λ-geodesically convex on the space of probability measures with fixed center of mass, a set which is preserved by the evolution equation, see [32, 18, 44, 2].

Remark 2.15 (Weak-L^p solutions)**.** Since weak measure solutions are equivalent to gradient flow solutions, our main uniqueness-stability Theorem 2.13 concludes the uniqueness of weak measure solutions to (1.2). Therefore, we can easily check that the previous constructed solutions in the series of papers [27, 7, 5, 20] for a family of more restrictive potentials W than the ones presented in this work, are indeed weak measure solutions up to their maximal time of existence. Let us make this statement more precise. It was shown in [8, Theorem 18] that weak-L^p solutions with initial data in $\mathcal{P}_2(\mathbb{R}^d)$ remain in $\mathcal{P}_2(\mathbb{R}^d)$ as long as they exist. These weak-L^p solutions satisfy equation (1.2) in the distributional sense, and they lead to curves in the space $\mathcal{P}_2(\mathbb{R}^d)\cap L^p(\mathbb{R}^d)$ which are continuous with respect to the strong topology in L^p, for suitable p, up to a maximal time of existence T^*. Hence, for potentials satisfying **(NL0)-(NL3)** and assuming the same additional conditions of growth at infinity in ∇W as in [27, 7, 5, 20, 8], one can show that the velocity field $\nabla W * \rho$ belongs to $L^1((0,T);L^2(\rho(t)))$ for all $0<T<T^*$, see for instance the proofs in [7, Section 3], [20, Section 2.2] or [8, Section 3]. Therefore, weak-L^p solutions with initial data in $\mathcal{P}_2(\mathbb{R}^d)$ are weak measure solutions up to their maximal time of existence, and thus, they do coincide up to that time with the weak measure solution constructed in Theorems 2.12-2.13. Let us also remark that in the works [27, 7, 5, 20, 8] the energy identity (2.39) was used as a tool for proving blow-up of the L^p-norm in finite time. To be more precise, for weak-L^p solutions one can prove an energy inequality like (2.39), where the equality sign is replaced by a less or equal sign, but the exact energy identity was missing. Hence our result in Theorem (2.12) also implies the energy identity for weak-L^p solutions.

Remark 2.16 (Comparison with classical PDE arguments)**.** Let us observe that a more classical strategy to construct weak measure solutions is based on approximating the initial datum by atomic measures, i.e. showing the convergence of the particle method. More precisely, one exploits the existence of solutions for the discrete particle system in

Remark 2.10 and the stability result in Theorem 2.13 to show convergence of the discrete approximating solutions to a limit curve. In this way, everything reduces to prove that the limit curve is a weak measure solution to (1.2), which is however not completely trivial, and would require some work. Moreover, it is not clear how to show directly that the weak measure solutions constructed in this way are both gradient flow solutions and curves of maximal slope, and that they satisfy the energy identity. This kind of strategy is well-known in kinetic theory, see for instance [22, 35, 40].

3. Particle measures in the JKO scheme

In this section, we show that the JKO scheme preserves the atomic part of the initial datum for all times, provided the time step is small enough. In particular, if we start with N-particles measure, it remains so, possibly with less particles, for all times. As a consequence, this immediately identifies the limit solution of the JKO scheme in this particular case. Moreover, it shows the well-posedness of a particle numerical scheme for solving numerically (1.2). More comments on this will be given below. Throughout this section, we allow λ to be positive, in which case our statements are stronger.

Given $\overline{\mu} \in \mathcal{P}_2(\mathbb{R}^d)$ let, for $\tau > 0$,

$$F_\tau[\mu] := \mathcal{W}[\mu] + \frac{1}{2\tau} d_W^2(\overline{\mu}, \mu). \tag{3.1}$$

Let us denote $u^- := \max\{0, -u\}$. We show that during a sufficiently small step of the JKO scheme, the mass contained in a particle remains concentrated, regardless of what the rest of the state looks like.

Definition 3.1 (Atomization). *Given $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, μ^* stands for the point mass located at the center of mass of μ, i.e:*

$$\mu^* := \delta_z \quad \text{where} \quad z = \int_{\mathbb{R}^d} x \, d\mu(x)$$

We say that μ^* is the atomization of μ.

Theorem 3.2. *Assume W satisfies* **(NL0)-(NL1)**. *Let $\overline{\mu} = m\delta_a + \mu_r \in \mathcal{P}_2(\mathbb{R}^d)$ with $0 < m \le 1$ and $\delta_a \perp \mu_r$. Given any $\tau > 0$ such that $\tau\lambda^- < 1$, let*

$$\mu \in \operatorname{argmin}_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} F_\tau[\nu], \tag{3.2}$$

and denote by π an optimal transportation plan between $\overline{\mu}$ and μ. Let us define

$$\mu_1(E) := \frac{1}{m} \pi(\{a\} \times E), \tag{3.3}$$

for any Borel set E. Then $\mu_1 = \mu_1^*$. In particular,

$$\mu = m\delta_z + \mu_s$$

for some $z \in \mathbb{R}^d$ and μ_s a nonnegative measure.

To rephrase the statement of the theorem in plain language: Any optimal transportation plan from the present state $\overline{\mu}$ to a minimizer of the JKO step μ carries all the mass from the particle at a to another point z. Thus the updated state has a particle at z, whose mass is at least the same as the one of the particle which was in a.

In case the measure $\overline{\mu}$ is a sum of N particles, by applying Theorem 3.2 to each particle, we easily conclude that μ is still a sum of particles, possibly less than N.

Corollary 3.3 (Particles remain particles)**.** *Assume W satisfies* **(NL0)-(NL1)**. *Let $\overline{\mu} = \sum_{i=1}^N m_i \delta_{x_i}$, where $x_1, \dots, x_N$ are distinct points in $\mathbb{R}^d$, $\sum_{i=1}^N m_i = 1$ and $m_i \in (0,1)$. Given any $\tau > 0$ such that $\tau\lambda^- < 1$, let*

$$\mu \in \operatorname{argmin}_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} F_\tau[\nu].$$

Then there exist $y_1, \dots, y_N \in \mathbb{R}^d$, not necessarily distinct, such that $\mu = \sum_{i=1}^N m_i \delta_{y_i}$.

To prove Theorem 3.2, given a minimizer μ of the JKO step, we show that

$$\mu_{new} := m\mu_1^* + (\mu - m\mu_1) \tag{3.4}$$

decreases the JKO functional:

$$F_\tau[\mu_{new}] < F_\tau[\mu], \qquad \text{if } \mu_1 \neq \mu_1^*. \tag{3.5}$$

This implies $\mu = \mu_{new}$. To prove (3.5) we examine what effect does atomizing μ_1 have on the two terms in the JKO functional: the energy and the Wasserstein distance. We show that, as expected, atomizing decreases the Wasserstein distance. On the other hand atomizing can increase the interaction energy, but only if λ is negative. The key observation is that in each of the terms the change is controlled by the variance of μ_1. Taking the time step small enough allows us to conclude.

Lemma 3.4. *Assume $\mathcal{W}$ satisfies* **(NL0)-(NL1)**. *Let $\nu_1, \nu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\nu = m_1\nu_1 + m_2\nu_2$ with $0 \le m_1 \le 1$ and $m_2 = 1 - m_1$. Let $\nu_{new} := m_1\nu_1^* + m_2\nu_2$. Then*

$$\mathcal{W}[\nu] - \mathcal{W}[\nu_{new}] \ge \frac{\lambda}{2} m_1 Var(\nu_1)$$

Proof. Introduce the symmetric bilinear form

$$B(\eta_1, \eta_2) = \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y)\, d\eta_1(x)\, d\eta_2(y)$$

so that

$$\begin{aligned}
\mathcal{W}[\nu] - \mathcal{W}[\nu_{new}] &= B(m_1\nu_1 + m_2\nu_2, m_1\nu_1 + m_2\nu_2) - B(m_1\nu_1^* + m_2\nu_2, m_1\nu_1^* + m_2\nu_2) \\
&= 2m_1m_2 B(\nu_1, \nu_2) + m_1^2 B(\nu_1, \nu_1) - 2m_1m_2 B(\nu_1^*, \nu_2) - m_1^2 B(\nu_1^*, \nu_1^*) \\
&= m_1 m_2 \int_{\mathbb{R}^d \times \mathbb{R}^d} [W(x-y) - W(z_1 - y)]\, d\nu_1(x)\, d\nu_2(y) \\
&\quad + \frac{m_1^2}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d} [W(x-y) - W(z_1 - z_1)]\, d\nu_1(x)\, d\nu_1(y),
\end{aligned}$$

where z_1 is the center of mass of ν_1. By the λ-convexity assumption **(NL1)**, for each $p \in \mathbb{R}^d$ and $r_p \in \partial W(p) \neq \emptyset$ the inequality

$$W(q) \ge W(p) + r_p \cdot (q-p) + \frac{\lambda}{2}|q-p|^2$$

holds for all $q \in \mathbb{R}^d$. Semiconvexity of W also implies that for $y \in \mathbb{R}^d$ there exists $r_{z_1 - y} \in \partial W(z_1 - y)$. Using this along with the fact that $0 \in \partial W(0)$ we obtain

$$\begin{aligned}
W(x-y) - W(z_1 - y) &\ge r_{z_1-y} \cdot (x - z_1) + \frac{\lambda}{2}|x - z_1|^2, \\
W(x-y) - W(0) &\ge \frac{\lambda}{2}|x-y|^2,
\end{aligned}$$

for all $x, y \in \mathbb{R}^d$. Since $\int_{\mathbb{R}^d}(x - z_1)\, d\nu_1(x) = 0$ we get

$$\begin{aligned}\mathcal{W}[\nu] - \mathcal{W}[\nu_{new}] &\geq m_1 m_2 \frac{\lambda}{2} Var(\nu_1) + \frac{m_1^2}{2}\frac{\lambda}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} |(x - z_1) - (y - z_1)|^2\, d\nu_1(x)\, d\nu_1(y)\\ &= m_1 m_2 \frac{\lambda}{2} Var(\nu_1) + m_1^2 \frac{\lambda}{2} Var(\nu_1)\\ &= \frac{m_1\lambda}{2} Var(\nu_1),\end{aligned}$$

which concludes the proof. □

Lemma 3.5. *Let $\overline{\mu}$ given as in Theorem 3.2. Given any $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, let π be an optimal transportation plan between $\overline{\mu}$ and ν and let ν_1 be defined by* (3.3)*. Let $\nu_{new} := m\nu_1^* + (\nu - m\nu_1)$. Then*

$$d_W^2(\overline{\mu}, \nu) - d_W^2(\overline{\mu}, \nu_{new}) \geq m\,\mathrm{Var}(\nu_1).$$

Proof. Let z be the center of mass of ν_1 (so that $\nu_1^* = \delta_z$). Denote by π_1 the restriction of π to $\{a\} \times \mathbb{R}^d$, and $\pi_2 := \pi - \pi_1$. Let $\pi_{new} := m\delta_{(a,z)} + \pi_2$. Note that π_{new} is a transportation plan between $\overline{\mu}$ and ν_{new}. Therefore,

$$\begin{aligned}d_W^2(\overline{\mu}, \nu) &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^2\, d(\pi_1 + \pi_2)\\ &= m\int_{\mathbb{R}^d} |y - a|^2\, d\nu_1(y) + \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^2\, d\pi_2\\ &= m\int_{\mathbb{R}^d} \left[|y - z|^2 + 2(y - z)\cdot(z - a) + |z - a|^2\right]\, d\nu_1(y) + \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^2\, d\pi_2\\ &= m\int_{\mathbb{R}^d} |y - z|^2\, d\nu_1(y) + \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^2\, d(m\delta_{(a,z)} + \pi_2)\\ &\geq mVar(\nu_1) + d_W^2(\overline{\mu}, \nu_{new}),\end{aligned}$$

as desired. □

Proof of Theorem 3.2. Assume that the claim does not hold, and consider μ_{new} defined by (3.4). Then, Lemmas 3.4 and 3.5 imply that

$$\begin{aligned}F_\tau[\mu] - F_\tau[\mu_{new}] &= \mathcal{W}[\mu] - \mathcal{W}[\mu_{new}] + \frac{1}{2\tau}\left(d_W^2(\overline{\mu}, \mu) - d_W^2(\overline{\mu}, \mu_{new})\right)\\ &\geq \frac{m}{2}\left(\lambda + \frac{1}{\tau}\right) Var(\mu_1) > 0,\end{aligned}$$

contradicting the minimality of μ. □

Remark 3.6. The above property of minimizers in each step of the JKO scheme carries over to the limiting solution, thanks to the convergence of the JKO scheme towards curves of maximal slope and gradient flow solutions of Section 2, see Theorem 2.8. Therefore, solutions corresponding to initial data with a finite number of particles plus an orthogonal part remain so for all times, with a possibly decreasing number of particles in time, see also Proposition 4.5. Moreover, combining Corollary 3.3 with the convergence of the JKO scheme in Theorem 2.8 allows to recover Remark 2.10, i.e. the correspondence between solutions of the ODE system (2.35) and gradient flow solutions of (1.2) with atomic initial measures.

4. Finite-time Total Collapse and Multiple Collapse by Stability

In this section, we focus on studying the large-time asymptotics of attractive non-Osgood potential, i.e. potentials satisfying assumptions **(NL0)-(NL3)** and **(NL-FTBU)**.

We start by discussing the monotonicity assumption in **(NL-FTBU)**. If the potential satisfies $w'(0^+) > 0$, i.e. it has a Lipschitz singularity at the origin, nearby particles move towards each other with a relative speed comparable to $w'(0^+)$, and thus we expect the concentration in finite time. In case **(NL-FTBUb)**, thanks to the non-Osgood condition, we do expect again concentration in finite time. In fact, in the case of a single particle subject to the potential $W(x)$, one easily checks that the particle reaches the origin in finite time. As we show in Theorem 4.3, compactly supported measures do collapse completely in finite time.

Remark 4.1 (No Oscillation Condition on the potential)**.** Let us point out that the condition of $w''(r)$ being monotone decreasing in a right interval of 0 is not too restrictive. Actually, plain monotonicity of w'' in a right interval of 0 together with the non-Osgood condition, the nonnegativity of w', and the fact that $w'(0^+) = 0$, imply that w'' is monotone decreasing on a right interval of 0 (the only possibility to violate this condition would be that the second derivative oscillate wildly at 0). To see this, note that the monotonicity of w'' implies that w' is either convex or concave in an interval $(0, \varepsilon_0)$. But w' cannot be convex, as otherwise its graph would be below the graph of the linear function $r \mapsto (w'(\epsilon_0)/\varepsilon_0)r$, which is not compatible with the integrability of $1/w'$ at 0.

Let us start by showing the finite total collapse in the case of finite number of particles.

Proposition 4.2 (Finite Time Particles Collapse)**.** *Assume W satisfies* **(NL0)-(NL3)** *and* **(NL-FTBU)***. Given the initial datum $\mu_0 = \sum_{i=1}^N m_i \delta_{x_i^0}$ with center of mass*

$$x_c := \sum_{i=1}^N m_i x_i^0,$$

let $\mu(t)$ denote the unique gradient flow solution with $\mu(0) = \mu_0$. Set R_0 to be the largest distance from the initial particles to the center of mass:

$$R_0 := \max_{i=1,\dots,N} |x_i^0 - x_c|.$$

Then there exists $T^* > 0$, depending only on R_0 but not on the number of particles, such that $\mu(t) = \delta_{x_c}$ for $t \geq T^*$.

Proof. Let us define the curves $t \mapsto x_i(t)$, $i = 1, \dots, N$ as the solution of the ODE system discussed in Remark 2.10:

$$\dot{x}_i = -\sum_{j \in C(i)} m_j \nabla W(x_i - x_j), \qquad i = 1, \dots, N$$

where $C(i) = \{j \in \{1, \dots, N\} \ : \ j \neq i, \, x_j(t) \neq x_i(t)\}$. Recall also that we define the sum over empty set of indexes to be zero. Then, $\mu(t) = \sum_{i=1}^N m_i \delta_{x_i(t)}$, where possibly $x_i(t) = x_j(t)$ for some $i \neq j$.

Our claim is equivalent to saying that there exists $T^* > 0$ such that $x_i(t) = x_c$ for all $t \geq T^*$ and $i = 1, \ldots, N$. Note that, due to assumption **(NL0)** the center of mass of the particles is preserved in time for the solutions of the ODE system. Since the system is translation invariant, we can assume that $x_c = 0$ without loss of generality.

We define the Lipschitz function $R(t)$ to be the distance of the furthest particle from the center of mass:

$$R(t) := \max_{i=1,\ldots,N} |x_i(t)|.$$

Recall that x_i are Lipschitz in time, and are C^1 for all but finitely many *collision times* $0 =: T_0 < T_1 < T_2 < \cdots < T_l < T_{l+1} := +\infty$.

We first compute a differential inequality for the function $R(t)$. Due to assumption **(NL-FTBU)**, for all $t \geq 0$ and all $i = 1, \ldots, N$

$$\begin{aligned} \frac{d^+ x_i}{dt} := \lim_{h \to 0^+} \frac{x_i(t+h) - x_i(t)}{h} &= - \sum_{j \in C(i)} m_j \nabla W(x_i - x_j) \\ &= - \sum_{j \in C(i)} m_j \frac{x_i - x_j}{|x_i - x_j|} w'(|x_i - x_j|). \end{aligned} \tag{4.1}$$

While it would have been sufficient to deal with the derivative $\frac{dx_i}{dt}$ which exists a.e., we wanted to highlight the fact that the right-hand derivative exists for all times, including the collision times. Using (4.1), we have

$$\begin{aligned} \frac{d^+}{dt} R^2(t) &= \max_{\{i \,:\, x_i(t) = R(t)\}} \frac{d^+}{dt} |x_i|^2 \\ &= \max_{\{i \,:\, x_i(t) = R(t)\}} -2 \sum_{j \in C(i)} m_j \frac{(x_i - x_j) \cdot x_i}{|x_i - x_j|} w'(|x_i - x_j|) \,. \end{aligned} \tag{4.2}$$

Note that since R is Lipschitz, $\frac{d}{dt} R^2$ exists a.e. and is equal to $\frac{d^+}{dt} R^2$. Observe that for any i as above, $(x_i - x_j) \cdot x_i \geq 0$ since all other particles are inside $\overline{B}(0, R(t))$. Using again assumption **(NL-FTBU)**, we have $w'(|x_i - x_j|) > 0$ and thus $\frac{d^+}{dt} R(t) \leq 0$, from which we deduce that $R(t) \leq R_0$ for all $t \geq 0$. Let us distinguish two cases:

Case (a): $w'(0^+) > 0$. Let us define

$$D := \min_{r \in [0, 2R_0]} w'(r) > 0.$$

By coming back to (4.2), we deduce that for all $t \geq 0$

$$\begin{aligned} \frac{d^+}{dt} R(t)^2 &\leq \max_{\{i \,:\, x_i(t) = R(t)\}} -2D \sum_{j \in C(i)} m_j \frac{(x_i - x_j) \cdot x_i}{|x_i - x_j|} \\ &\leq \max_{\{i \,:\, x_i(t) = R(t)\}} -\frac{D}{R(t)} \sum_{j \neq i} m_j (x_i - x_j) \cdot x_i \end{aligned}$$

since $|x_i - x_j| \leq 2R(t)$ for $j \neq i$ and $(x_i - x_j) \cdot x_i \geq 0$. It is easy to check, using the unit total mass of the measure and that the center of mass is zero, that for any i as above

$$\sum_{j \neq i} m_j (x_i - x_j) \cdot x_i = R(t)^2.$$

Hence $\frac{d^+}{dt}R(t) \le -D$. We conclude that the claim holds with $T^* = R_0/D$.

Case (b): $w'(0^+) = 0$ together with the other assumptions in **(NL-FTBUb)**. The function w' is then concave on $(0, \varepsilon_0)$. Together with the fact that $w'(0^+) = 0$ this implies that $w'(r)/r$ is decreasing in $(0, \varepsilon_0)$. Without of loss of generality, we can assume that $\varepsilon_0 < \varepsilon_1$, with ε_1 as in **(NL-FTBUb)**.

Let us first show that $R(t)$ must reach values less than $\varepsilon_0/2$ in finite time. Since $R(t)$ is decreasing, it suffices to consider the case $R_0 > \varepsilon_0/2$. Fix any time such that $R(t) \ge \varepsilon_0/2$. Coming back to (4.2), we distinguish for any i such that $R(t) = |x_i(t)|$, two sets of particles: I, where $|x_i - x_j| \le \varepsilon_0/2$, and II, where $|x_i - x_j| > \varepsilon_0/2$. For indexes in the set I we can use that $w'(r)/r$ is decreasing, while to handle the set II we define

$$D := \min_{r \in [\varepsilon_0/2, 2R_0]} w'(r) > 0.$$

Using again $|x_i - x_j| \le 2R(t)$ for $j \ne i$ and $(x_i - x_j) \cdot x_i \ge 0$, we can write

$$\frac{d^+}{dt}R^2(t) \le \max_{\{i\,:\,x_i(t)=R(t)\}} \left\{ -2\frac{w'(\varepsilon_0)}{\varepsilon_0} \sum_{(I)} m_j(x_i - x_j) \cdot x_i - \frac{D}{R(t)} \sum_{(II)} m_j(x_i - x_j) \cdot x_i \right\}.$$

Thanks to $R(t) \ge \varepsilon_0/2$ and $w'(\varepsilon_0) \ge D$, we can finally conclude that

$$\frac{d^+}{dt}R^2(t) \le \max_{\{i\,:\,x_i(t)=R(t)\}} -\frac{D}{R(t)} \sum_{j \ne i} m_j(x_i - x_j) \cdot x_i = -DR(t)$$

for all times such that $R(t) \ge \varepsilon_0/2$. Thus, there exists a time τ such that $R(t) \le \varepsilon_0/2$ for $t \ge \tau$. We now refine the above argument for $t \ge \tau$ using that the distance between any two particles satisfies $|x_i - x_j| \le 2R(t) \le \varepsilon_0$. Since $w'(r)/r$ is decreasing on $(0, \varepsilon_0)$ we deduce that for times $t \ge \tau$

$$\frac{d^+}{dt}R(t)^2 \le \max_{\{i\,:\,x_i(t)=R(t)\}} -\frac{w'(2R(t))}{R(t)} \sum_{j \ne i} m_j(x_i - x_j) \cdot x_i = -w'(2R(t))R(t),$$

so that $\frac{d}{dt}R(t) \le -w'(2R(t))/2$ for almost all $t \ge \tau$. Using the non-Osgood condition, i.e. the integrability of $1/w'(r)$ at the origin, we conclude that $R(t) = 0$ for a certain T^* completely determined by the inequality $\frac{d}{dt}R(t) \le -w'(2R(t))/2$.

Let us remark that this proof shows that the time of total collapse of the particles to their center of mass does not depend either on the number of particles or their masses, but only on R_0. □

Making use of the stability result, the convergence of the particle method, and the total collapse for finite number of particles, we deduce the second main result of this work.

Theorem 4.3 (Finite Time Total Collapse)**.** *Assume W satisfies* **(NL0)-(NL3)** *and* **(NL-FTBU)***. Let $\mu(t)$ denote the unique gradient flow solution starting from the probability measure μ_0 with center of mass*

$$x_c := \int_{\mathbb{R}^d} x \, d\mu_0,$$

supported in $\overline{B}(x_c, R_0)$. Then there exists T^*, depending only on R_0, such that $\mu(t) = \delta_{x_c}$ for all $t \ge T^*$.

Proof. As in the previous proposition, we can assume $x_c = 0$. Given any compactly supported measure μ_0 in $B(0, R_0)$ and any $\eta > 0$, we can find a number of particles $N = N(\eta)$, a set of positions $\{x_1^0, \dots, x_N^0\} \subset B(0, R_0)$, and masses $\{m_1, \dots, m_N\}$, such that

$$d_W\left(\mu_0, \sum_{i=1}^N m_i \delta_{x_i^0}\right) \leq \eta.$$

Let us denote by $\mu_\eta(t)$ the particle solution associated to the initial datum $\mu_\eta(0) = \sum_{i=1}^N m_i \delta_{x_i^0}$.

By Proposition 4.2, there exists a time T^* independent of N such that $\mu_\eta(t) = \delta_0$ for $t \geq T^*$. Hence, by the stability result in Theorem 2.13 we obtain

$$d_W(\mu(T^*), \delta_0) = d_W(\mu(T^*), \mu_\eta(T^*)) \leq e^{-\lambda T^*} d_W\Big(\mu_0, \sum_{i=1}^N m_i \delta_{x_i^0}\Big) \leq e^{-\lambda T^*}\eta.$$

By the arbitrariness of η, we conclude that $\mu(t) = \delta_0$ for all $t \geq T^*$ as desired. □

Remark 4.4 (Finite Time Total Collapse and Tail Behavior)**.** The previous result can be generalized for measures which are not compactly supported by the following procedure. Let us consider the case in which $c_0 := \inf_{[0,+\infty)} w' = w'(0^+) > 0$. Then the proof of case (a) in Proposition 4.2 shows that, if μ_0 is supported in $B(x_c, R_0)$, then $\mu(t) = \delta_{x_c}$ for $t \geq R_0/c_0$. From this fact and the stability estimate, it is not difficult to show that for any initial datum μ_0 decaying more than exponentially at infinity (say a gaussian), $\mu(t)$ converges exponentially fast to δ_{x_c} in infinite time. Indeed, if

$$\mu_{0,R} := \frac{\mu_0 \lfloor_{B(x_c,R)}}{\mu_0(B(x_c, R))},$$

then one easily gets $d_W(\mu_0, \mu_{0,R}) \lesssim e^{-\overline{C}R}$ for any $\overline{C} > 0$, and their centers of mass x_c and $x_{c,R}$ are exponentially close too. Hence, if $\mu_R(t)$ denotes the solution starting from $\mu_{0,R}$, then $\mu_R(R/c_0) = \delta_{x_{c,R}}$. Therefore, choosing $\overline{C} > 2|\lambda|/c_0$, we get

$$d_W(\mu(t), \delta_{x_c}) \leq d_W(\mu(t), \mu_{c_0 t}(t)) + |x_c - x_{c,c_0 t}| \lesssim e^{-\lambda t} d_W(\mu_0, \mu_{0,c_0 t}) + e^{-\overline{C}c_0 t} \lesssim e^{-\overline{C}c_0 t},$$

as desired. As expected the tail behavior of the initial measure has to be fast enough to compensate the exponential growing bound in the stability when $\lambda < 0$. On the other hand, if $\lambda \geq 0$ then we do not need any assumption on the initial datum to prove convergence in infinite time, although having estimates on the tails allows to prove better rates of convergence.

The aim of the following proposition is to show that, if we start with a measure which has some atomic part, then the atoms can only increase their mass. We present a proof based on particle approximations, an alternative approach is using the JKO-scheme, via Theorem 3.2.

Proposition 4.5 (Dirac Deltas can only increase mass)**.** *Let $\mu(t)$ denote the unique solution starting from the probability measure $\sum_{i=1}^N m_i \delta_{x_i^0} + \nu_0$, and define the curves $t \mapsto x_i(t)$, $i = 1, \dots, N$, as the solution of the ODE*

$$\dot{x}_i(t) = -(\partial^0 W * \mu(t))(x_i(t)).$$

Then $\mu(t) \geq \sum_{i=1}^N m_i \delta_{x_i(t)}$ for all $t \geq 0$, with possibly $x_i(t) = x_j(t)$ for some $t > 0$, $i \neq j$.

Proof. This result is again an application of the result in the case of a finite number of particles, combined with the stability of solutions. Let us approximate ν_0 with a sequence $\nu_0^k = m\sum_{j=1}^k \frac{1}{k}\delta_{y_j}$, with $m := \nu_0(\mathbb{R}^d)$. Then the unique solution starting from $\sum_{i=1}^N m_i\delta_{x_i^0} + \nu_0^k$ is given by

$$\mu^k(t) = \sum_{i=1}^N m_i \delta_{x_i^k(t)} + \sum_{j=1}^k \frac{m}{k}\delta_{y_j^k(t)},$$

where $x_i^k(t)$ and $y_j^k(t)$ solve the ODE system

$$\dot{x}_i^k = -\sum_{l\neq i,\, x_l^k \neq x_i^k} m_j \nabla W(x_l^k - x_i^k) - \sum_{l,\, y_l^k\neq x_i^k} \frac{m}{k}\nabla W(y_l^k - x_i^k), \qquad i = 1,\dots,N,$$

$$\dot{y}_j^k(t) = -\sum_{i,\, x_i^k\neq y_l^k} m_i \nabla W(x_i^k - y_j^k) - \sum_{l\neq j,\, y_l^k \neq y_j^k} \frac{m}{k}\nabla W(y_l^k - y_j^k), \qquad j = 1,\dots,k.$$

This gives in particular

$$\mu^k(t) \geq \sum_{i=1}^k m_i \delta_{x_i^k(t)} \tag{4.3}$$

as measures, since the particles coming from the "discretization" of ν_0 can only join the fixed particles x_i but they will not split them. We now observe that the curves $t \mapsto x_i^k(t)$ are uniformly Lipschitz (locally in time). Indeed to obtain a bound on the velocity $\partial^0 W * \mu$, by (1.5) it suffices to show that the second moments of the measures $\mu^k(t)$ are uniformly bounded, locally in time. To check this, we use as test function $|x|^2$ for a general gradient flow solution $\mu(t)$ of (1.2), and exploiting the λ-convexity of W we get

$$\begin{aligned}
\frac{d}{dt}\int_{\mathbb{R}^d} |x|^2\, d\mu(t)(x) &= -2\int_{\mathbb{R}^d} x\cdot(\partial^0 W * \mu(t))\, d\mu(t)(x)\\
&= -\int_{x\neq y} (x-y)\cdot(\nabla W(x) - \nabla W(y))\, d\mu(t)(y)\, d\mu(t)(x)\\
&\leq -\lambda \int_{x\neq y} |x-y|^2\, d\mu(t)(y)\, d\mu(t)(x) \leq 4|\lambda| \int_{\mathbb{R}^d} |x|^2\, d\mu(t)(x).
\end{aligned}$$

Therefore, using the stability of solutions and Ascoli-Arzelà Theorem, up to a subsequence each curve $t\mapsto x_i^k(t)$ converges locally uniformly to a limit curve $t \mapsto x_i(t)$ which satisfies

$$\dot{x}_i(t) = -\int_{\mathbb{R}^d} \partial^0 W(y - x_i(t))\, d\mu(t)(x), \qquad i = 1,\dots,N.$$

Taking the limit in the inequality (4.3) we get the desired result. □

Finally, let us show that the blow up of the L^∞ norm of a solution to (1.2) may occur at a time strictly less than the time of total collapse. In order to produce such a phenomenon, we shall work again with the ODE system (2.35) and then we argue by approximation. We first show a simple argument in a particular situation. We introduce some notation following Proposition 4.2. Let us define the curves $t \mapsto x_i(t)$, $i = 1, \dots, 2N$ as the solution of the ODE system

$$\frac{dx_i}{dt} = -\sum_{j \in C(i)} \frac{1}{2N} \nabla W(x_i - x_j), \qquad x_i(0) = x_i^0, \qquad i = 1, \dots, 2N,$$

so that $\mu(t) = \sum_{i=1}^{2N} \frac{1}{2N} \delta_{x_i(t)}$ is a gradient flow solution to (1.2). We define $x_{c_1}(t)$ and $x_{c_2}(t)$ to be the center of masses of the first N and the last N particles respectively. Let us consider the functions

$$R_1(t) := \max_{i=1,\dots,N} |x_i(t) - x_{c_1}(t)| \qquad \text{and} \qquad R_2(t) := \max_{i=N+1,\dots,2N} |x_i(t) - x_{c_2}(t)|,$$

and denote by $\mu_1(t)$ and $\mu_2(t)$ the measures $\sum_{i=1}^{N} \frac{1}{2N} \delta_{x_i(t)}$ and $\sum_{i=N+1}^{2N} \frac{1}{2N} \delta_{x_i(t)}$ respectively.

Proposition 4.6 (Multiple Collapse)**.** *Assume the potential W satisfies* **(NL0)-(NL3)**, **(NL-FTBUa)**, *and* $\lim_{x \to +\infty} w'(x) = 0$. *There exist $r_0, d_0, T_0, T_1 > 0$ such that if* $\max\{R_1(0), R_2(0)\} \le r_0$ *and* $|x_{c_1}(0) - x_{c_2}(0)| \ge d_0$*, then*

$$\mu_1(t) = \delta_{x_{c_1}(t)} \ne \mu_2(t) = \delta_{x_{c_2}(t)} \qquad \textit{for all } T_0 \le t < T_1.$$

Proof. The ODE system satisfied by the particles is given by

$$\frac{dx_i}{dt} = -\sum_{j \in C(i)} \frac{1}{2N} \frac{x_i - x_j}{|x_i - x_j|} w'(|x_i - x_j|) \,, \qquad i = 1, \dots, 2N.$$

We distinguish two sets of particles: (I) the set of the first N particles and (II) the set of last N particles. Arguing as in Proposition 4.2, we obtain

$$\frac{d^+}{dt} R_1^2(t) = \max_{\{i \,:\, |x_i(t) - x_{c_1}(t)| = R_1(t)\}} -\sum_{j \in C(i)} \frac{1}{N} \frac{(x_i - x_j) \cdot (x_i - x_{c_1})}{|x_i - x_j|} w'(|x_i - x_j|) \,.$$

with $(x_i - x_j) \cdot (x_i - x_{c_1}) \ge 0$ for $j = 1, \dots, N$. Fix d_0 large enough such that $|w'(r)| \le \frac{1}{4} w'(0^+)$ for $r \ge d_0/2$. Then, as long as $\max\{R_1(t), R_2(t)\} \le \frac{1}{8} d_0$ and $|x_{c_1}(t) - x_{c_2}(t)| \ge \frac{3}{4} d_0$, we have that for some i for which $|x_i(t) - x_{c_1}(t)| = R_1(t)$

$$\begin{aligned} \frac{d^+}{dt} R_1^2(t) &\le -\frac{w'(0^+)}{2N\, R_1(t)} \sum_{(I)} (x_i - x_j) \cdot (x_i - x_{c_1}) + \frac{|x_i - x_{c_1}|}{N} \sum_{(II)} \big| w'(|x_i - x_j|) \big| \\ &\le -\frac{w'(0^+)}{2} R_1(t) + \frac{w'(0^+)}{4} R_1(t) = -\frac{w'(0^+)}{4} R_1(t), \end{aligned} \tag{4.4}$$

where we used

$$\sum_{(I)} (x_i - x_j) \cdot (x_i - x_{c_1}) = N R_1(t)^2.$$

By continuity in time of solutions, there exists $t_* > 0$ small enough such that $|x_{c_1}(t) - x_{c_2}(t)| \ge \frac{3}{4} d_0$ is satisfied for $0 \le t \le t_*$. Choosing $r_0 \le \min\{\frac{1}{8} d_0, \frac{w'(0^+)}{8} t_*\}$ and using (4.4), we ensure that $\max\{R_1(t), R_2(t)\} \le \frac{1}{8} d_0$ in $0 \le t \le t_*$ and $R_1(t_*) = 0$. Analogously, we have that $R_2(t_*) = 0$. Then, it is clear by continuity in time that we can choose $T_0 \le t_* < T_1$ such that the statement holds. ☐

By a more refined analysis, one could produce an analogous result in case the potential W satisfies **(NL0)-(NL3)**, **(NL-FTBUb)**, and $\lim_{x \to +\infty} w'(x) = 0$. For instance, one can explicitly construct examples of particle configurations with special symmetries where one can check by tedious computations the multiple collapse phenomena. As a consequence, we obtain the following result

Corollary 4.7. *Assume the potential W satisfies* **(NL0)**-**(NL3)**, **(NL-FTBU)**, *and* $\lim_{x\to+\infty} w'(x) = 0$. *Then, there exists a nonnegative function $\rho_0 \in C_c^\infty(\mathbb{R}^d)$ with unit mass and there two curves $x_{c_k}(t)$, $k = 1, 2$, and $0 < T_0 < T_1$ such that the gradient flow solution associated with the initial datum $\rho_0\, dx$ satisfies*

$$\mu(t) = \frac{1}{2}\delta_{x_{c_1}(t)} + \frac{1}{2}\delta_{x_{c_2}(t)} \quad \text{and} \quad x_{c_1}(t) \neq x_{c_2}(t)$$

for all $t \in (T_0, T_1)$.

It is clear from the previous proof that this two particle collapse can be generalized to multiple collapse situations with as many particle collapses as we want and choosing the time ordering of their collapses in any desired manner.

Acknowledgements. JAC acknowledges support from the project MTM2008-06349-C03-03 DGI-MCI (Spain) and 2009-SGR-345 from AGAUR-Generalitat de Catalunya. The work of MDF is partially supported by Award No. KUK-I1-007-43, made by King Abdullah University of Science and Technology (KAUST), CRM-Barcelona, the DAMTP - University of Cambridge and the 2007 Azioni Integrate Italia-Spagna 25. DS is grateful to NSF for support via the grants DMS-0638481 and DMS-0908415, and is also thankful to the CNA (NSF grant DMS-0635983) for its support during the preparation of this paper. All authors acknowledge IPAM-UCLA where this work was started during the thematic program on "Optimal Transport". The authors are also thankful to the referees for a number of valuable suggestions that helped to improve the paper.

References

[1] M. Agueh, *Existence of solutions to degenerate parabolic equations via the Monge-Kantorovich theory*, Adv. Differential Equations, 10, No. 3 (2005), 309-360.

[2] L. Ambrosio, N. Gigli and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, Lectures in Mathematics, Birkhäuser, (2005).

[3] L. Ambrosio and G. Savaré, *Gradient Flows of Probability Measures* in Handbook of Evolution Equations Dafermos, C. and Feireisl, E. (ed.); Vol. 3, Elsevier (2006), 1-136.

[4] D. Benedetto, E. Caglioti, M. Pulvirenti, *A kinetic equation for granular media*, RAIRO Modél. Math. Anal. Numér., 31, (1997), 615–641.

[5] A.L. Bertozzi and J. Brandman, *Finite-time blow-up of L^∞-weak solutions of an aggregation equation*, to appear in Comm. Math. Sci.

[6] A.L. Bertozzi, J.A. Carrillo and T. Laurent, *Blowup in multidimensional aggregation equations with mildly singular interaction kernels*, Nonlinearity, 22 (2009), 683-710.

[7] A.L. Bertozzi and T. Laurent, *Finite-time Blow-up of Solutions of an Aggregation Equation in* $\mathbb{R}^n$, Comm. Math. Phys., 274 (2007), 717-735.

[8] A.L. Bertozzi, T. Laurent and J. Rosado, *L^p Theory for the multidimensional aggregation equation*, to appear in Comm. Pure Appl. Math.

[9] P. Biler, G. Karch, P. Laurençot, *Blowup of solutions to a diffusive aggregation model*, preprint (2009).

[10] A. Blanchet, V. Calvez and J.A. Carrillo, *Convergence of the mass-transport steepest descent scheme for the subcritical Patlak-Keller-Segel model*, SIAM J. Numer. Anal., 46, (2008), 691-721.

[11] A. Blanchet, J.A. Carrillo, and N. Masmoudi, *Infinite Time Aggregation for the Critical Patlak-Keller-Segel model in* $\mathbb{R}^2$, Comm. Pure Appl. Math., 61, (2008), 1449-1481.

[12] A. Blanchet, J. Dolbeault, and B. Perthame, *Two-dimensional Keller-Segel model: optimal critical mass and qualitative properties of the solutions*, Electron. J. Differential Equations, (2006), No. 44, 32 pp. (electronic).
[13] M. Bodnar and J.J.L. Velázquez, *An integro-differential equation arising as a limit of individual cell-based models*, J. Differential Equations, 222, (2006), 341-380.
[14] S. Boi, V. Capasso and D. Morale, *Modeling the aggregative behavior of ants of the species Polyergus rufescens*, Spatial heterogeneity in ecological models (Alcalá de Henares, 1998), Nonlinear Anal. Real World Appl., 1, (2000), 163-176.
[15] M. Burger, V. Capasso and D. Morale, *On an aggregation model with long and short range interactions*, Nonlinear Analysis. Real World Applications. An International Multidisciplinary Journal, 8 (2007), 939-958.
[16] M. Burger and M. Di Francesco, *Large time behavior of nonlocal aggregation models with nonlinear diffusion*, Networks and Heterogeneous Media, 3, (2008), 749-785.
[17] J.A. Carrillo, M.R. D'Orsogna and V. Panferov, *Double milling in self-propelled swarms from kinetic theory*, preprint UAB (2008).
[18] J.A. Carrillo, R.J. McCann and C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates*, Rev. Matemática Iberoamericana, 19, (2003), 1-48.
[19] J.A. Carrillo, R.J. McCann and C. Villani, *Contractions in the 2-Wasserstein length space and thermalization of granular media*, Arch. Rat. Mech. Anal., 179, (2006), 217–263.
[20] J.A. Carrillo and J. Rosado, *Uniqueness of Bounded Solutions to Aggregation Equations by Optimal Transport Methods*, to appear in the proceedings of the 5th-European Congress of Mathematicians (2008).
[21] Y.-L. Chuang, Y.R. Huang, M.R. D'Orsogna and A.L. Bertozzi, *Multi-vehicle flocking: scalability of cooperative control algorithms using pairwise potentials*, IEEE International Conference on Robotics and Automation, (2007), 2292-2299.
[22] R. Dobrushin, *Vlasov equations*, Funct. Anal. Appl., 13 (1979), 115-123.
[23] F. Golse, *The mean-field limit for the dynamics of large particle systems*, Journées "Équations aux Dérivées Partielles", Exp. No. IX, 47, Univ. Nantes, Nantes, (2003).
[24] Y. Huang and A. L. Bertozzi, *Self-similar blow-up solutions to an aggregation equation in* $\mathbb{R}^n$, preprint (2009).
[25] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal., 29, (1998), 1-17.
[26] E.F. Keller and L.A. Segel, *Initiation of slide mold aggregation viewed as an instability*, J. Theor. Biol., 26, (1970).
[27] T. Laurent, *Local and Global Existence for an Aggregation Equation*, Comm. Partial Differential Equations, 32, (2007), 1941-1964.
[28] H. Li and G. Toscani, *Long–time asymptotics of kinetic models of granular flows*, Arch. Rat. Mech. Anal., 172, (2004), 407-428.
[29] D. Li and J. Rodrigo, *Finite-time singularities of an aggregation equation in* $\mathbb{R}^n$ *with fractional dissipation,* to appear in Comm. Math. Phys.
[30] D. Li and J. Rodrigo, *Refined blowup criteria and nonsymmetric blowup of an aggregation equation,* to appear in Adv. Math.
[31] D. Li and X. Zhang, *On a nonlocal aggregation model with nonlinear diffusion,* preprint (2008).
[32] R.J. McCann, *A convexity principle for interacting gases*, Adv. Math., 128, (1997), 153-179.
[33] A. Mogilner and L. Edelstein-Keshet, *A non-local model for a swarm*, J. Math. Bio., 38, (1999), 534-570.
[34] D. Morale, V. Capasso and K. Oelschläger, *An interacting particle system modelling aggregation behavior: from individuals to populations*, J. Math. Biol., 50, (2005), 49-66.
[35] H. Neunzert, *The Vlasov equation as a limit of Hamiltonian classical mechanical systems of interacting particles*, Trans. Fluid Dynamics, 18, (1977), 663-678.
[36] A. Okubo, S. Levin, *Diffusion and Ecological Problems: Modern Perspectives*, Springer, Berlin, (2002).
[37] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations, 26, (2001), 101-174.

[38] C.S. Patlak, *Random walk with persistence and external bias*, Bull. Math. Biophys., 15, (1953), 311-338.
[39] F. Poupaud, *Diagonal defect measures, adhesion dynamics and Euler equation*, Methods Appl. Anal., 9, (2002), 533-561.
[40] H. Spohn, *Kinetic equations from Hamiltonian dynamics: Markovian limits*, Rev. Modern Phys., 52, (1980), 569-615.
[41] C.M. Topaz and A.L. Bertozzi, *Swarming patterns in a two-dimensional kinematic model for biological groups*, SIAM J. Appl. Math., 65, (2004), 152-174.
[42] C.M. Topaz, A.L. Bertozzi, and M.A. Lewis, *A nonlocal continuum model for biological aggregation*, Bulletin of Mathematical Biology, 68, No. 7 (2006), 1601-1623.
[43] G. Toscani, *One-dimensional kinetic models of granular flows*, RAIRO Modél. Math. Anal. Numér., 34, No. 6 (2000), 1277-1291.
[44] C. Villani, *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI, (2003).
[45] C. Villani, *Optimal transport, old and new*, Grundlehren der mathematischen Wissenschaften 338, Springer, (2009).

[1] ICREA and Departament de Matemàtiques, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain. E-mail: carrillo@mat.uab.es.

[2] Sezione di Matematica per l'Ingegneria, Dipartimento di Matematica Pura ed Applicata, Università di L'Aquila, Piazzale E. Pontieri 2, Monteluco di Roio, 67040 L'Aquila, Italy. E-mail: difrance@univaq.it.

[3] Department of Mathematics, The University of Texas at Austin, 1 University Station C1200, Austin, TX 78712, USA. E-mail: figalli@math.utexas.edu.

[4] Department of Mathematics, University of California - Los Angeles, Los Angeles, CA 90095, USA. E-mail: laurent@math.ucla.edu.

[5] Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, USA. E-mail: slepcev@math.cmu.edu.